372
٨.٩.٨٠

Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

VOLUME 69, 1962

BUREAU OF EDNL. & PSYL.

HELEN ORR
Managing Editor

ELIZABETH S. REED
Advertising Manager

MARY SUSAN GILLETT
Editorial Assistant

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

Editorial Consultants

Joe K. Adams
Robert P. Abelson
Gordon W. Allport
Justin M. Aronfreed
Solomon E. Asch
Albert Bandura
S. Howard Bartley
Daniel E. Berlyne
Dalbir Bindra
Abraham H. Black
Edgar F. Borgatta
Everett W. Bovard
Gordon H. Bower
Joseph V. Brady
Jack W. Brehm
Donald M. Broverman
Roger W. Brown
Josef Brozek
F. Robert Brush
Robert R. Bush
John B. Carroll
Raymond B. Cattell
David H. Grant
Richard Christie
Russell M. Church
Sidney E. Cleveland
Clyde H. Coombs
Lee J. Cronbach
Karl M. Dallenbach
James Deese
William N. Dember
Carl P. Duncan
J. A. Easterbrook
Ward Edwards
James P. Egan
Charles W. Eriksen
William K. Estes
Seymour Feshbach
Leon Festinger
Edwin A. Fleishman
Edward Forgy
John Gaito
Wendell R. Garner
Frank A. Geldard
James J. Gibson
Kenneth P. Goodrich
Albert E. Goss
Clarence H. Graham
David A. Grant
G. Robert Grice
J. P. Guilford
Harold Gulliksen
Norman Guttman
Calvin S. Hall
D. O. Hebb
R. J. Herrnstein
Eckhard H. Hess
Winfred F. Hill
Herbert Hochberg
Julian Hochberg
Lloyd G. Humphreys
J. McV. Hunt
Leo M. Hurvich
Ray Hyman
Francis W. Irwin
James J. Jenkins
Lyle V. Jones
Michael Kahn
Leon J. Kamin
Howard H. Kendler
Tracy S. Kendler
William Kessen
Leonard S. Kogan
Wolfgang Kohler
Raymond G. Kuhlen
William W. Lambert
Douglas Howard Lawrence
Daniel S. Lehrman
Marvin Levine
J. C. R. Licklider
Gardner Lindzey
Frank A. Logan
Ardie Lubin
R. Duncan Luce
David T. Lykken
Craig MacAndrew
Kenneth MacCorquodale
William J. McGill
F. Joseph McGuigan
Donald W. MacKinnon
Robert B. MacLeod
Irving Maltzman
George Mandler
Sarnoff A. Mednick
Samuel J. Messick
Walter Mischel
O. Hobart Mowrer
Conrad G. Mueller
Bennet B. Murdock
Clyde E. Noble
Charles E. Osgood
Helen Peak
Albert Pepitone
Charles C. Perkins, Jr.
Carl Pfaffmann
Leo J. Postman
Malcolm G. Preston
Victor Raimy
Gregory Razran
K. Edward Renner
Frank Restle
Lorrin A. Riggs
Irvin Rock
Carl R. Rogers
Milton E. Rosenbaum
Frank Rosenblatt
David L. Rosenhan
Mark R. Rosenzweig
Burton S. Rosner
William W. Rozeboom
George E. Ruff
Seymour B. Sarason
Stanley Schachter
Harold Schlosberg
John P. Seward
Roger N. Shepard
Joseph Shoben
Sidney Siegel
Herbert A. Simon
Karl U. Smith
M. Brewster Smith
Moncrieff H. Smith
Kenneth W. Spence
Roger W. Sperry
Arthur W. Staats
Larry Stein
Eliot Stellar
Saul H. Sternberg
Joseph C. Stevens
Patrick Suppes
John A. Swets
Renato Tagiuri
Philip Teitelbaum
Han-Lukas Teuber
Warren S. Torgerson
Leona E. Tyler
Benton J. Underwood
William S. Verplanck
Victor H. Vroom
Edward L. Walker
Seymour Wapner
Robert W. White
Delows D. Wickens
Morton Wiener
David R. Williams
Julius Wishner
David Zeaman

LANCASTER PRESS, INC., LANCASTER, PA.

Edn.

CONTENTS OF VOLUME 69

Amsel, Abram. Frustrative Nonreward in Partial Reinforcement and Discrimination Learning: Some Recent History and a Theoretical Extension 306
Anderson, Norman H. On the Quantification of Miller's Conflict Theory 400
Becker, Gordon M. Objective Measures of Subjective Probability and Utility 136
Berger, Seymour M. Conditioning through Vicarious Instigation 450
Boneau, C. Alan. A Comparison of the Power of the *U* and *t* Tests 246
Bower, Gordon H. An Association Model for Response and Training Variables in Paired-Associate Learning 34
Briggs, Michael H., and G. Barrie Kitto. The Molecular Basis of Memory and Learning 537
Broverman, Donald M. Normative and Ipsative Measurement in Psychology 295
Champion, R. A. Stimulus-Intensity Effects in Response Evocation 428
Cornsweet, Tom N. Changes in the Appearance of Stimuli of Very High Luminance .. 257
Davis, James H. See Restle, Frank
Deese, James. On the Structure of Associative Meaning 161
Deutsch, J. A. A System for Shape Recognition 492
Edwards, Ward. Subjective Probabilities Inferred from Decisions 109
Epstein, Alan N. See Teitelbaum, Philip.
Fringke, Gerald. See Johnson, Ronald C.
Gagné, Robert M. The Acquisition of Knowledge 355
Gibson, James J. Observations on Active Touch 477
Grant, David A. Testing the Null Hypothesis and the Strategy and Tactics of Investigating Theoretical Models 54
Gulick, W. L. See Smith W. M.
Henle, Mary. On the Relation between Logic and Thinking 366
Johnson, Ronald C. Reanalysis of "Meaningfulness and Verbal Learning" 233
Johnson, Ronald C., Thomson, Calvin W., and Fringke, Gerald. A Response to Stricker's Comments on "Word Values, Word Frequency, and Visual Duration Thresholds" 239
Jones, Joan E. All-or-None Versus Incremental Learning 156
Jones, Joan E. Contiguity and Reinforcement in Relation of CS-UCS Intervals in Classical Aversive Conditioning 176
Jones, Marshall B. Practice as a Process of Simplification 274
Kendler, Howard H., and Kendler, Tracy S. Vertical and Horizontal Processes in Problem Solving 1
Kendler, Tracy S. See Kendler, Howard H.
Kenshalo, D. R., and Nafe, John P. A Quantitative Theory of Feeling: 1960 17
Kitto, G. Barrie. See Briggs, Michael H.
Logan, Frank A. Conditional-Outcome Choice Behavior in Rats 467
Luce, R. Duncan. Comments on Rozeboom's Criticism of "On the Possible Psychophysical Laws" 548
Mandler, George. From Association to Structure 415
Mednick, Sarnoff A. The Associative Basis of the Creative Process 220
Nafe, John P. See Kenshalo, D. R.
Peabody, Dean. Two Components in Bipolar Scales: Direction and Extremeness 65

PRUITT, DEAN G. Pattern and Level of Risk in Gambling Decisions 187
PUBOLS, JR., BENJAMIN H. An Application of Levine's Model for Hypothesis Behavior to Serial Reversal Learning 241
RESTLE, FRANK. The Selection of Strategies in Cue Learning 329
RESTLE, FRANK, AND DAVIS, JAMES H. Success and Speed of Problem Solving by Individuals and Groups 520
ROZEBOOM, WILLIAM W. Comment 552
ROZEBOOM, WILLIAM W. The Untenability of Luce's Principle 542
RUBIN, LEONARD S. Patterns of Adrenergic-Cholinergic Imbalance in the Functional Psychoses 501
SCHACHTER, STANLEY, AND SINGER, JEROME E. Cognitive, Social, and Physiological Determinants of Emotional State 379
SINGER, JEROME E. See Schachter, Stanley.
SMITH, W. M., AND GULICK, W. L. A Statistical Theory of Dynamic Contour Perception 91
SOLOMON, RICHARD L., AND TURNER, LUCILLE H. Discriminative Classical Conditioning in Dogs Paralyzed by Curare Can Later Control Discriminative Avoidance Responses in the Normal State 202
SUNDLAND, DONALD M. The Construction of *Q* Sorts: A Criticism 62
TEITELBAUM, PHILIP, AND EPSTEIN, ALAN N. The Lateral Hypothalamic Syndrome: Recovery of Feeding and Drinking after Lateral Hypothalamic Lesions 74
THOMSON, CALVIN W. See Johnson, Ronald C.
TULVING, ENDEL. Subjective Organization in Free Recall of "Unrelated" Words 344
TURNER, LUCILLE H. See Solomon, Richard L.
WHITE, SHELDON H. A Note on Intensity Generalization and Prothetic Scaling 149

VOL. 69, NO. 1 JANUARY 1962

PSYCHOLOGICAL REVIEW

VERTICAL AND HORIZONTAL PROCESSES IN PROBLEM SOLVING[1]

HOWARD H. KENDLER
New York University
AND
TRACY S. KENDLER
Barnard College

The present paper is concerned with *an* approach—and not *the* approach—to the universally appealing but nevertheless unpopular research area of problem solving. Problems of problem solving have proved to be particularly refractory to psychologists. More often than not the uncommon researcher with the temerity to attack some aspect of reasoning retreats to more secure and conventional problems when he discovers that his sorties fail to achieve any impressive victory. As a result the literature of problem solving is almost chaotic because it is so heavily sprinkled with isolated bits of information (Duncan, 1959).

Perhaps the present stage of development of psychology does not justify the strategy of investigating such a complex phenomenon. Fortunately, or not, science has no built-in traffic lights to inform investigators when to proceed. It may be a risky and potentially unfruitful gambit to investigate problem solving but then again it may not be. In addition to the intrinsic interest of the area it does offer a challenge to those psychologists who are interested in testing the generality of any set of theoretical principles stemming from other areas of behavior (e.g., learning, perception).

This paper initially will make fleeting references to some methodological problems with which a researcher in the field of reasoning must contend. Then a simple pretheoretical model of problem solving will be described, followed by a report of research which the model generated, and which in turn is shaping the model itself.

METHODOLOGICAL PROBLEMS IN PROBLEM SOLVING RESEARCH

Anybody who does research is—or should be—aware that every decision he makes cannot be justified by facts or logic. Some decisions must be made on the basis of personal intuition. This is particularly true for the researcher in problem solving who must make three strategic decisions which cannot help but have profound influences on his research and the ideas they generate (Kendler, 1961). These decisions, which are not completely independent, are related to the place of problem solving in psychology, the use of complex or simple experimental tasks, and the selection of a pretheoretical model to guide research. Considering the volitional nature of these problems, as well as the current status of psychological knowledge, it would be both inappropriate and erroneous to consider these methodological problems as offering only one sensible alternative. Adopting this point of view would do much to minimize the needless disputation that seems to perennially surround matters of research strategy.

---

[1] An earlier version of this paper was delivered by Howard H. Kendler as an invited address to the 1960 meeting of the Eastern Psychological Association, which was held in New York City. The authors are indebted to the Office of Naval Research and the National Science Foundation for their support of the research reported in this paper.

Accepting the principle that a basic research strategy is not simply an outgrowth of logical and factual considerations does not reduce one to making decisions in either a haphazard or random manner. A given strategy can be adopted on the basis of rational considerations as long as it is realized that other reasonable attitudes might lead to the adoption of different decisions.

The history of problem solving in particular and psychology in general suggests that problem solving can best be conceptualized not as a basic psychological process, but instead as one that reflects the interaction of more fundamental processes (e.g., learning, perception, and motivation).

If problem solving is not viewed as a unitary process, how is an appropriate experimental situation selected to investigate it? One possibility is that a problem can be selected from a "true life" situation such as troubleshooting electronic equipment. Or problems can be invented (Duncker, 1945; Maier, 1930) that capture the flavor, if only partially, of problems we meet in everyday life.

A more analytical approach can be taken to the selection of an experimental situation to investigate problem solving. If problem solving is compounded of elementary behavioral processes, then it may be more strategic to devise some simple problems in which the relationships of fundamental psychological mechanisms to problem solving are highlighted. That is, tasks should be devised not to duplicate or imitate everyday problems, but instead to isolate and magnify the basic mechanisms that operate in such complex tasks.

This analytical approach which is favored by the authors suffers from one major drawback. How is it possible to know the basic mechanisms of problem solving prior to their discovery? Obviously, excepting divination, there is no method. But this does not prevent the analytical approach from operating. The researcher can prejudge theoretical issues by formulating a model of what he guesses problem solving to be like. The model can guide the investigator in selecting the hypotheses to test, as well as the experimental situations in which to test them.

This brings us to the third and most important decision a problem solving researcher has to make: his choice of a pretheoretical model (Koch, 1959). A pretheoretical model is not equivalent to a theory. The criterion of validity cannot properly be applied to it because essentially a pretheoretical model is an informal conception that operates as an analogy (Lachman, 1960). It is conceivable that different models (e.g., learning, perception, information theory) can all lead to fruitful and valid theories of problem solving.

Psychologists have many possibilities from which to choose their model. These models can be conveniently divided into two main categories: the empirical model that springs primarily from experimental data, and the formal model that is usually generated by mathematical or logical systems. Among the empirical models that have achieved some acceptance are those that are based on introspective findings (e.g., the four successive stage model of "preparation," "incubation," "inspiration," and finally "verification"), the facts of perception, and those of learning. Some formal models used are those dependent upon stochastic models, game theory, and the operation of computers.

The present authors adopted an S-R learning pretheoretical model. The decision no doubt was influenced by professional training and past research efforts. But other considerations entered. For the past 4 decades S-R learning psychologists have probably been the most active experimental and theoretical group in psychology. To some, if not a large, extent this can be attributed to the fruitful and cleansing effect S-R language has upon designing, reporting, and interpreting research. S-R language forces the psychologist to focus his attention on objectively defined environmental and behavior variables and thus encourages the collection of data and the testing of ideas. The efforts of S-R learning psychologists have supplied a host of facts, concepts, and hypotheses that can be exploited in an exploratory excursion into the realm of problem solving.

The facts and theories of learning, however, do not spontaneously coalesce to form a model that can guide research in problem solving. Some selection must be made. S-R learning theory does not represent a single organized formulation. Anyone who is familiar with the systematic orientations of Hull (1952), Guthrie (1952), Spence (1956), and Skinner (1953) is aware of this. Many of these systematic differences, however, become attenuated and some even disappear when viewed from the distance of problem solving behavior. It is possible and perhaps even profitable to develop a learning model for problem solving that ignores many of the points of disagreement among S-R theories.

Much of the objection to S-R language stems from the apparent discrepancy between active, flowing behavior and the inert, static, single S-R association. Using S-R language does not mean that complex behavior *actually* consists of S-R connections. After analyzing the concept of light, Toulmin (1953), concludes: "We do not *find* light atomized into individual rays: we *represent* it as consisting of such rays" (p. 29). Applying the same idea to the concept of the S-R association: "We do not *find* behavior atomized into individual S-R associations: we *represent* it as consisting of such S-R associations." The concept of the S-R association, therefore, must be judged not in terms of its ability to provide a clear image of behavior, but rather in its capacity to represent the facts of behavior.

PRETHEORETICAL MODEL OF PROBLEM SOLVING

An S-R model needs to represent two important characteristics of problem solving behavior. These characteristics are behavior is continuous, and at any one time behavior consists of several habits. The terms "horizontal" and "vertical" are used to refer to these processes; horizontal to the continuity of behavior against the dimension of time, and vertical to the assumption that independent levels of behavior (i.e., S-R units) occur simultaneously.

The assumption that S-R associations do not occur in isolation, but instead are linked together to form integrated, continuous behavior goes back many years (e.g., Watson, 1913). Today the process is most commonly referred to as chaining. Skinner (1953) and his associates have developed powerful techniques that shape

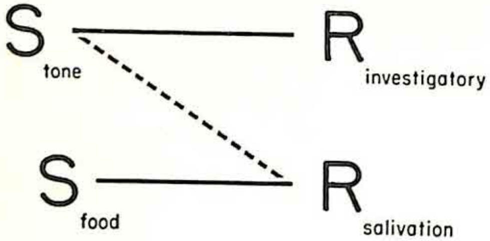


FIG. 1. An S-R representation of classical conditioning.

behavior into long, complicated chains. The mass of data they have collected suggests important principles governing habit chaining. There is litle doubt that when their quasitheoretical system is exploited fully with autoinstructional devices that important insights into problem solving behavior will emerge, particularly in relation to how an added bit of knowledge can trigger problem solution. The kind of chaining with which the Skinnerians have dealt (i.e., adding new S-R units to an already functioning chain) does not exhaust all the problems associated with the horizontal processes of problem solving. Of particular importance to problem solving is the *spontaneous* integration of separate habits which occurs when an organism infers the consequences of combining previously independent S-R units. This kind of chaining was investigated in Kohler's (1925) classical studies of insight and in the more controlled reasoning experiments of Maier (1930). More recently the authors (Kendler & Kendler, 1956, 1961; Kendler, Kendler, Pliskoff, & D'Amato, 1958) have tried to identify some of the important variables that enable children to combine separate experiences in order to solve an inference-type problem. Much of the research reported in this paper will be concerned with how mediated stimulus and response events aid in the formation of problem solving chains.

The assumption of vertical processes, i.e., the organism responds several different ways at any one time, is also not a novel one. Every psychologist is aware that organisms make several different responses simultaneously, although typically only one is attended to. Sometimes the different responses are interrelated, as is the case between the heart and respiration rates of a fearful organism. In other cases the different responses are independent, e.g., a person's conversation is uninfluenced by his tugging at his ear lobe. The best laboratory example of vertical processes, and one that has much relevance to problem solving, is shown in Figure 1. Those familiar with introductory psychology textbooks will recognize this diagram as representing classical conditioning. Notice that the two solid lines indicate independent S-R units which are operating simultaneously. One is the tone that initiates the "investigatory" response, and the other is the food which elicits salivation. Initially these two associations operate in a *parallel* fashion, but as a result of their simultaneous occurrence an *interaction* takes place which is expressed by the broken line representing the acquired conditioned response.

Obviously the brief reference to horizontal and vertical processes in which it is assumed fundamental S-R principles operate (e.g., discrimination, generalization, etc.) presents at best the barest skeleton of a model of problem solving. It needs the flesh and skin of experimental facts to give it solidity and theoretical principles to clothe it in scientific respectability. Let us now review some of the progress that has been made in this direction.

CONCEPT LEARNING AND UTILIZATION

Although the primitive model just described fails to generate any research by itself, it does suggest that individual experiments cannot be directed at *problem solving in its entirety.* There are too many aspects to this phenomenon. The researcher, in designing an experiment, must scan the entire problem solving process and then focus upon that segment that promises to yield fruitful results and is also amenable to investigation.

For reasons that will become evident, it was decided to compare reversal and nonreversal shifts in a simple concept learning task. Figure 2 characterizes each kind of shift by showing a *simplified* version of an experimental situation used with children. The stimuli (cups) for their first discrimination differed simultaneously on two dimensions (size and brightness). The subject is rewarded for responses to one dimension (e.g., large cup is positive, small cup is negative). The other dimension is irrelevant. After learning the first discrimination, the subject is forced to shift to another response. In a reversal shift the subject is required to respond to the same dimension on which he was originally trained, but his overt choice has to be reversed, e.g., he has to shift from a *large* cup to a *small* one. For a nonreversal shift the previously irrelevant dimension becomes relevant, e.g., black becomes positive after large had been positive.

Buss (1953) reported that college students executed a reversal shift more rapidly than a nonreversal shift. He attributed this superiority to the intermittent reinforcements that retard the progress of a nonreversal shift. For example, in Figure 2,[2] when a subject is making a nonreversal shift from large positive to black positive, he is reinforced when choosing the large black cup in preference to the small white cup. This fortuitous reinforcement of the choice of the large cup helps maintain the size discrimination and hence retards the learning of the brightness discrimination. The reversal shift group, on the other hand, receives no reinforcement of the previously correct responses, since they are 100% *non*reinforced.

This analysis is at best incomplete. The work of Kendler and Vineberg (1954) suggested that adult human concept learning cannot be represented adequately by a single-unit S-R theory in which the external stimulus is directly connected to the overt response. Instead, a mediational mechanism (see Figure 3) is required which assumes that the external stimulus evokes an implicit response which produces an implicit cue that is connected to the overt response.

It would be useful to digress for a

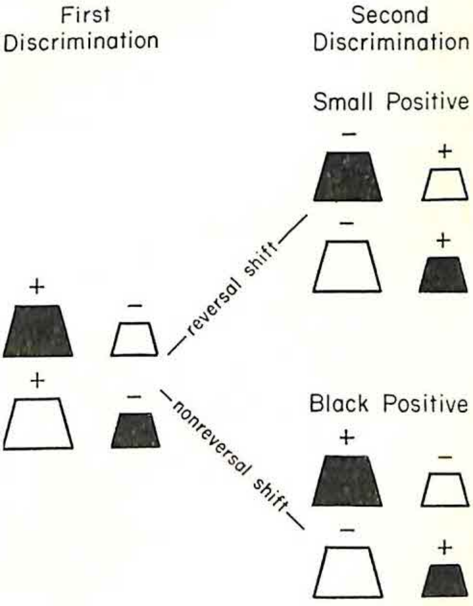


FIG. 2. Examples of a reversal and a nonreversal shift.

---

[2] The purpose of Figure 2 is to clarify the meaning of both a reversal and nonreversal shift. It would be misleading to believe that it represents *exactly* the methodology of "reversal-nonreversal" studies reported in this paper. For all experiments reported, except that of Buss (1953), designs were used that controlled for fortuitous intermittent reinforcements effects in a nonreversal shift.

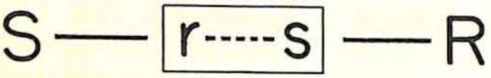


FIG. 3. A schematic representation of the mediational hypothesis.

moment to comment about the epistemological status of these inferred stimulus and response events which are enclosed in the rectangle to emphasize their hypothetical character. Although not directly observable, they are "tied to" environmental and behavioral events. The basic assumption of the mediational hypothesis, at least for the time being, is that the implicit stimulus and response events obey the same principles that operate in observable S-R relationships.

The mediational hypothesis has generated confusion. Perhaps the following brief statements will clarify some possible areas of misunderstanding.

1. The mediational hypothesis is neither new nor revoluntionary. Meyer (1911) and Watson (1913) referred to it, and Hull (1930) gave it a more formal status by coining the concept of the "pure stimulus act." Guthrie (1952) has always laid heavy stress on a mediational-type hypothesis when emphasizing the importance of proprioceptive stimulation in learning.

2. The implicit stimulus and response events *need not* be conceived as having an existence independent of their relation to independent and dependent variables. These implicit events are theoretical constructs. Their epistemological status is closer to such concepts as drive and habit than to directly observable stimulus and response events.

Some mediating events can conceivably and probably will be coordinated to introspective reports, language behavior, muscular movements, and other observable events. Coordinations of this sort can be useful in developing mediational theory. But such coordinations are not *essential* to mediational theory. The fact that genes are not

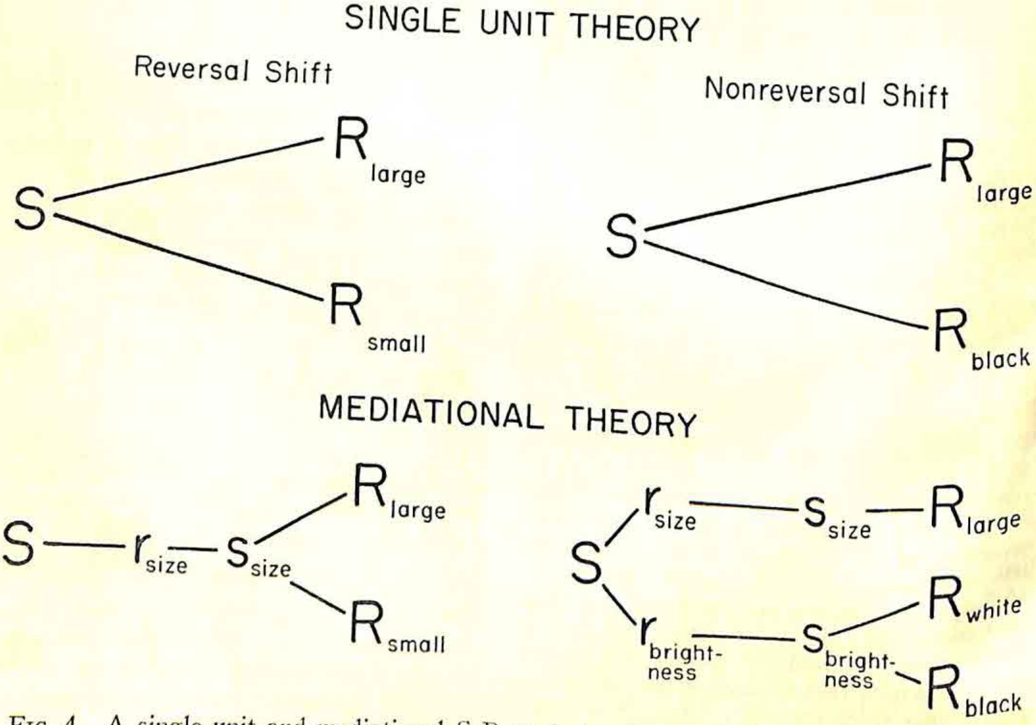


FIG. 4. A single unit and mediational S-R analysis of a reversal and nonreversal shift.

directly observable (at least according to the geneticists consulted) does not interfere with their theoretical and practical usefulness. Even if it were possible to observe a gene directly, it would be necessary to distinguish between it as an observable entity and as a concept within a nomological network. It would be unwise, and strategically shortsighted, to *identify* mediational events with introspective reports or language behavior, or other observable events. The "validity" of the mediational mechanism does not depend on being coordinated with observable events, but depends instead on being utilized in a successful explanatory system.

Figure 4 characterizes reversal and nonreversal shifts in terms of both a single unit S-R analysis and a mediational one.[3] It would be predicted, according to a single unit hypothesis, that if fortuitous intermittent reinforcements were eliminated from a nonreversal shift, it would occur more rapidly than a reversal shift. The reason for this is that at the time of the shift the difference between the strength of the dominant incorrect habit and the to-be-correct habit is much greater for the reversal, as compared to the nonreversal shift. Consequently more training will be required to make the correct habit dominant in a reversal shift. According to the mediational theory the situation is entirely different. A reversal shift enables the subject to utilize the same mediated response. Only the overt response has to be changed. A nonreversal shift, on the other hand, required the acquisition of a *new* mediated response, the cues of which have to be attached to a *new* overt response. Because the old mediational sequence has to be discarded and a new one formed, the nonreversal shift should be executed more slowly than a reversal shift.[4] Thus, if it were possible to eliminate fortuitous intermittent reinforcements, then the stage would be set for a crucial experiment testing the conflicting implications of the single-unit and mediational S-R theories. The results of a series of such crucial experiments (Buss, 1956; Harrow & Friedman, 1958; Kendler & D'Amato, 1955) have been consistent with the mediational formulation in showing that college students execute a reversal shift more rapidly than a nonreversal shift. It is im-

---

[3] Figure 4 highlights the problem of what are the effective stimuli that are associated to the overt response in both a reversal and nonreversal shift. It is not intended to be a detailed analysis of which there may be several alternatives. For example, in a single unit theory the habit to choose the large container might result from learning two separate specific habits (e.g., the choice of a large black container when coupled with a small white one and the selection of a large white container when paired with a small black one). Another possibility, which would be consistent with Spence's theory (1936), is that the response is to the effective stimulus *large* since responses to the other features of the environment are not consistently reinforced. Similarly adult subjects in a reversal shift might use the mediator *size* or *large* or both. The effective stimulus which is controlling the organism's response must be determined by experimentation. The point made here is that the general implications of the single unit and mediational theories, as discussed in this paper, would be the same for a number of different effective stimuli.

[4] There are two possible ways of analyzing the superiority of a reversal shift over a nonreversal shift within an S–R mediational framework. One is to simply count the number of new associations that have to be formed. As Figure 4 indicates only one new association has to be formed in a reversal shift while two have to be formed for a nonreversal shift. Another possibility is that a mediating response is more difficult to extinguish than is an overt response. For the present the formulation can remain open-ended until information relevant to these two alternatives is gathered.

TABLE 1

MEAN NUMBER OF TRIALS TO CRITERION ON TEST DISCRIMINATION FOR SUBJECTS SCORING ABOVE AND BELOW THE MEDIAN ON THE TRAINING DISCRIMINATION

Group	Performance on training discrimination	
	Above Median (slow learners)	Below Median (fast learners)
Reversal	24.4	6.0
Nonreversal	9.0	15.5

portant to note that in a similar kind of problem rats find a nonreversal shift easier than a reversal shift (Kelleher, 1956). Thus, one is forced to conclude that a single unit S-R theory accurately represents the behavior of rats, while mediational S-R theory is required for the concept learning of articulate humans.

The discontinuity between the behavior of rats and college students directs one's attention toward the conditions responsible for the development of mediational processes. Somewhere on a hypothetical evolutionary dimension between the rat and college student there should be a point where a transition is made from a single unit to mediational control. An obvious place to locate this point would be in the behavior of young children.

A study with kindergarten children (Kendler & Kendler, 1959) showed that these children as a group executed a reversal and nonreversal shift at approximately the same rate. One might conclude that the point in human development was discovered which was psychologically halfway between the white rat and the college student, since the kindergarten children were neither responding in a single unit nor mediational manner, but instead in some compromise fashion. Another possibility is that the children had reached a transitional stage in development, in which the task to which they were subjected led some to function on a single unit basis, and others to operate with a mediational mechanism. If half of the subjects respond in each way, the total results would have revealed no difference between the two kinds of shifts.

The second alternative seems to fit the data. When the kindergarten children were divided into fast and slow learners on the basis of their performance in the first problem (training discrimination), slow learners performed during the second problem (test discrimination) according to the single unit theory; like rats they found a nonreversal shift easier. Fast learners, on the other hand, performed in accordance with the mediational theory; like college students, they found a reversal shift easier. These results were interpreted as demonstrating that these kindergartners, taken as a group, were in the process of developing mediating responses relevant to this task, and that some were further along than others.

If this interpretation be correct, then it would follow that for a group of younger (i.e., preschool) children a still smaller proportion should develop appropriate mediating responses. It

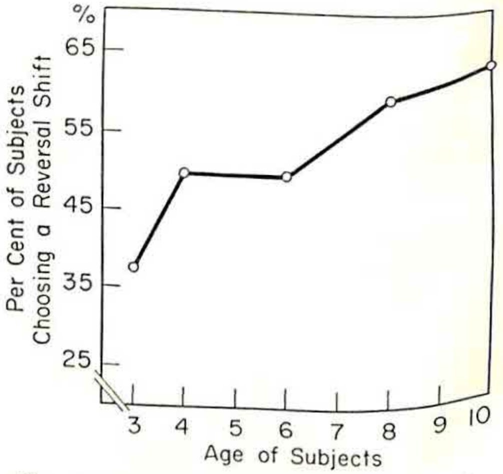


FIG. 5. Percentage of children responding in a reversal shift manner as a function of age.

would be expected that such a group, taken as a whole, would show clearcut evidence of the superiority of a nonreversal over a reversal shift. An experiment (Kendler, Kendler, & Wells, 1960) designed to test this hypothesis produced results consistent with this prediction; like rats, nursery school children found a nonreversal shift to be easier than a reversal shift.

In a very recent study the experimental procedure was modified so that after learning the initial discrimination, the children of 3, 4, 6, 8, and 10 years of age who served as subjects, had a choice of either responding in a reversal or a nonreversal manner. Under such circumstances, it would be expected that the proportion of children who respond in a reversal manner would increase with age. Figure 5 shows that the percentage of children who chose a reversal shift rose gradually from 37.5 at 3 to 62.5 at 10.

Generalizing from all of these results, it would seem that in their early development, children tend to respond in a manner consistent with a single unit S-R theory. With age, they develop a tendency to respond in a mediational manner. The last study cited suggests that it is, or will soon be, possible to ascertain the lawful relationship governing the course of this development.

The point of these experiments is not to classify children into one of two categories: rat-like or human-like. Their aim is to lay the groundwork for experiments designed to investigate the mediational process itself. If one wants to investigate mediational processes, does it not seem sensible to scrutinize them at the time when they are developing? Answering this question in the affirmative, it was decided to investigate the relationship between the hypothesized mediational processes and verbal behavior—a relationship everybody assumes to be intimate and important.

Particularly relevant to this attempt to coordinate verbalization with mediation were observations that during the course of the experiments just described, it was not uncommon for children to verbalize spontaneously the correct solution while simultaneously making an incorrect choice. A few children did this for many consecutive trials. This observation is relevant to the concept of vertical processes. Two chains of habits are occurring simultaneously. One has to do with verbal response; the other with the overt choice. For these children the two chains are parallel, that is, they do not interact.

Luria (1957), the Russian psychologist, made somewhat similar observations in his research with children. He explains this sort of phenomenon in the following way:

> In the early stages of child development, speech is only a means of communication with adults and other children. . . . Subsequently it becomes also a means whereby he organizes his own experience and regulates his own actions. So the child's activity is mediated through words (p. 116).

These observations and their interpretations of noninteracting parallel processes point to the complex interrelationships existing between verbal behavior on the one hand and problem solving on the other. If nothing else, they destroy the illusion that it is reasonable to describe an organism as verbal or nonverbal without considering the problem with which it is confronted. The terms verbal and nonverbal become meaningful—and fruitful—when related to specific problem solving tasks.

It would seem fruitful to investigate the cue function of words for children of two age levels. One possibility is that age influences problem solving

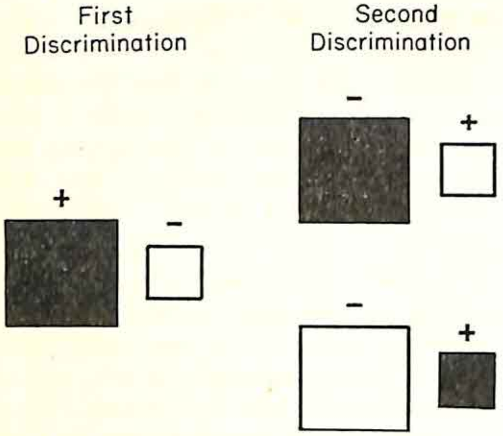


FIG. 6. The experimental procedure used to study the influence of verbal habits on a reversal shift.

only in so far as it leads to the acquisition of words. If younger children, say 4 years of age, could acquire the same words as 7-year-olds, they would solve a simple concept-learning problem the same way. The other possibility is that the acquisition of the verbal label by itself is not sufficient; the word must be integrated with other behavioral chains to influence problem solving behavior. And for this to happen some developmental changes must first take place.

In order to test these two alternatives, children of 4 and 7 years of age were presented with another variation of the reversal shift problem as shown in Figure 6. They initially learned a simple discrimination between a pair of stimuli that varied simultaneously in size and brightness. In the illustration provided in Figure 6, the large black square is correct. While they were learning, the children were required to verbalize aloud the stimuli to which they were responding. One-third learned to say "large" (or "small" as the case may be) by the simple device of instructing them to tell the experimenter which was correct, the large or the small one. Another third learned to say "black" (or "white") in a corresponding way. The remaining third was not required to say anything. After learning the discrimination, all subjects were presented with a reversal shift. In the example depicted in Figure 6, the shift is to small regardless of size. Thus, the group that initially described the correct stimulus as "large" had verbalized the relevant dimension. The verbal response of "black" was irrelevant to this reversal shift.

Figure 7 shows the results of the three experimental groups for the two age levels. If developmental processes affect the utilization of verbal responses in problem solving, then it would be expected that the three verbalization conditions (which produced a significant main effect) would influence the behavior of the two age groups differently. These results suggest, but not quite at a significant level, that there is an interaction effect. Figure 7 shows that the younger children profited by making the kind of verbal response appropriate to a reversal shift, while they were hindered by learning

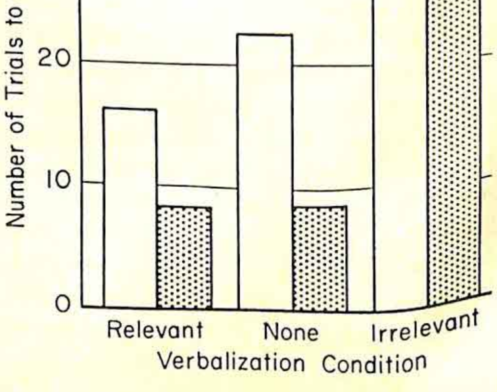


FIG. 7. The effect of verbalizations on a reversal shift for 4- and 7-year-old children.

inappropriate verbal responses. With no verbalization the 7-year-old children who presumably were responding largely in a mediational manner, accomplished a reversal shift much more rapidly than their younger counterparts. But unlike the 4-year-olds, they did not profit from being trained to make the relevant responses. At 7 years of age they are capable of making the response themselves and outside help appears to be of little use. In contrast, the influence of irrelevant verbalizations is marked. The performance of the 7-year-olds was even poorer than that of the 4-year-olds, suggesting that the interfering effects of being given an inappropriate mediated response are greater when one is capable of spontaneously generating the correct one (7-year-olds) than when one is not (4-year-olds).

How are these data to be explained? Attributing differences to developmental factors is not sufficient. It is necessary to represent developmental differences in terms of the concepts of the behavior model that is being used. That is, if a verbal label for a young child does not possess the same cue function as it does for an older child, then it becomes necessary to specify how and why this comes about. To some extent this has been done by emphasizing the transition from a single unit to a mediational system, as well as suggesting that with age an increase occurs in interaction among chains of different vertical levels. But obviously this analysis of the developmental process demands further theoretical and empirical development.

These studies are intimately related to the oft-reported finding that many species of subhuman animals are able to make a fairly rapid reversal shift *if* they receive a previous series of such shifts. Rats (Buytendijk, 1930; Krechevsky, 1932) show a marked improvement in executing successive reversals. They finally reach a point (Dufort, Guttman, & Kimble, 1954), in a **T** maze, where they learn to go to a new rewarded goal after making only one error. Even more dramatic are the rapid discrimination reversals exhibited by Harlow's (1949) monkeys. But fish (Wodinsky & Bitterman, 1957) exhibit only a slight improvement in successive reversals, while isopods (invertebrates) show no improvement (Thompson, 1957).

Because of the necessity to use somewhat different experimental procedures for different species, it is difficult to draw an unqualified conclusion about the ability of different species to transfer what has been learned from previous reversal shifts to a new one. But the suggestion is strong that as you ascend the evolutionary scale organisms acquire a greater capacity to generate cues that enable them to make rapid reversal shifts. This behavior, according to our analysis, borders on the language responses of humans. The main difference is that our human subjects, except those of a very young age, exhibit rapid reversals without any previous reversal training. Whereas the human automatically seems to generate a mediated response that provides the basis for his rapid reversal, the animal subject must gradually acquire an ability to respond appropriately to some response produced cue resulting from nonreinforcement of a previously correct response.

Up to now, the reversal and nonreversal technique has been used to investigate mediational and developmental variables. It has proved sufficiently flexible to be used in a study (Kendler, Glucksberg, & Keston, 1961) which was designed to lengthen a problem solving chain so that the interaction between various segments could be observed. In this study a

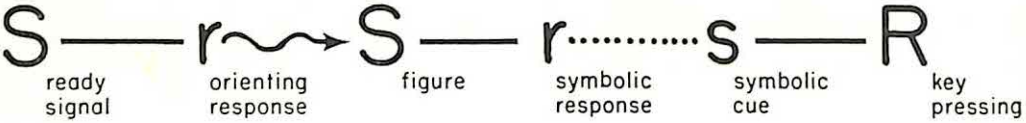


FIG. 8. The hypothesized behavioral chain operating at the time the subject was being shifted to the second concept. (Capital letters refer to directly observable stimulus and response events, while small letters refer to those that are inferred.)

perceptual orienting S-R unit was added on to the mediational chain already described. Figure 8 illustrates in an oversimplified manner the behavioral sequence involved in this study in which subjects had to learn to press the correct button when two physically discrete and spatially separate stimulus patterns were projected on a screen at such a rapid rate that only one could be perceived on any trial. During the learning of each of two successive concepts (involving either a reversal or nonreversal shift), the subject had to pay attention to the relevant stimulus pattern while ignoring the irrelevant one. Thus, in order to make the correct overt response consistently, a subject initially had to make the appropriate orienting response in order to perceive the relevant stimulus pattern to which he had to make the correct mediational response which served as the cue for the key-pressing act.

An experimental design was used in which, at the time of the shift from the first to the second concept, one group had already learned the appropriate orienting response as well as the appropriate mediating act. They needed only to learn a new terminal key-pressing response. The shift, for them, was easy to make. In contrast, the behavior of three other experimental groups was significantly worse. One group had to learn a new orienting response, e.g., look to the left instead of the right. Another group had to learn a new mediated response (i.e., they were required to make a nonreversal shift). The last group had to acquire both a new orienting and mediated response. The fact that the groups which were missing one or both of the necessary behavior units (orienting and mediated responses) did not differ significantly among themselves, as well as being much poorer than the group that had both, highlights the problem of synchronizing the S-R units in a behavioral chain. The advantage in this study of having one appropriate unit without the other is at best negligible. The reason for this is that reinforcement is only achieved consistently when both the appropriate orienting and mediating responses are operating. This particular study points to the need for discovering laws associated with the strengthening and weakening of independent S-R units in a problem solving chain, as well as the principles governing their synchronization.

This study also highlighted a very basic problem in all of these reversal studies. This problem has to do with the very first correct response following the reversal shift. After discovering that the previous mode of responding is erroneous, what makes the subject change his response, i.e., push the button that was previously wrong? Introspective reports fail to provide any clearcut answer and even if they did they would be in need of explaining (Kendler, 1961).

One hypothesis is that the selection of the new correct response is due to the operation of a behavioral chain in addition to the one described in Figure

8. The first nonreinforcement in a reversal shift sets off a chain, the consequence of which is to select the response other than the one that was previously correct. This may result from a number of different reasons (e.g., logical considerations, forgetting, etc.). The important point, however, is that the new key-sorting response occurs contiguously with the implicit mediational response appropriate to a reversal shift. As a result, a new association is formed between the old implicit cue and the new key-pressing response.

In essence, what is being stated is that adult subjects, when making or deciding to make the first correct postshift response, do not adopt the *principle* underlying a reversal shift. Instead, it is assumed processes are operating which encourage the selection of the correct response while an implicit cue appropriate to a reversal shift is operating. This sort of an analysis was described previously (Kendler & Mayzner, 1956) as

> sort of a James-Lange theory of problem solving . . . one makes the overt correct . . . response and if the appropriate symbolic cue is present, then problem solution will occur (pp. 247–248).

Guthrie (1959) says the same thing more neatly: *"What is being noticed becomes a signal for what is being done"* (p. 186).

Again the authors would like to guard against giving the impression of oversimplifying a terribly complex problem. They do not believe the contiguous occurrence of an implicit cue from one chain with the correct overt response from another tells the whole story. This new association in order to persist must be reinforced and in some manner "fit into" the subject's ongoing behavioral chains.

The emphasis on this vertical connection between a cue and a response from different chains is related in a distant way to Hebb's (1958) stressing the role of "chance" in problem solving:

> There are few scientists who have not had the experience of setting out to solve problem A and ending up instead with the solution to B. . . . This is serendipity, the art of finding one thing while looking for another (p. 215).

According to the present analysis, serendipity results from the adventitious and contiguous occurrence of a cue and a response which are themselves segments from different behavior chains. Theoretically it should be possible to demonstrate this point experimentally by training subjects to respond simultaneously to two separate tasks. A problem then would have to be presented that requires for its solution the combination of a stimulus from one chain with the response from the other. In such an experimental situation, controlling the time relationship between the two should have an important effect on problem solving. Presumably contiguity between the two should provide the most optimal conditions for problem solving (Underwood, 1952). The development of this kind of experimental procedure should allow for parametric studies of the basic variables of the phenomenon which has commonly been called "insight," as well as throw light upon issues raised by others (e.g., Cofer, 1957; Maltzman, 1955; Saugstad, 1957).

The pretheoretical model that guides the present research has many more facets that can be exploited. Only one will now be mentioned. Glucksberg (1962), for example, extended neobehavioristic drive theory (Spence, 1956) to problem solving. He used a functional-fixedness problem (Adamson, 1952, Duncker, 1945) in which the correct response in the habit hierarchy could either be made to be low or high. If the correct habit was low, it would be expected that a strong drive would retard problem solving because it would retard the extinction of the dominant incorrect response (Kendler & Lachman, 1958; Perin, 1942). Since drive energizes behavior, a high drive should facilitate problem solving performance when the correct habit is dominant. The findings were consistent with this analysis.

Because functional-fixedness problems are often represented in perceptual terms, Glucksberg was interested in seeing whether the same drive model could be applied to a simple perceptual recognition problem in which subjects were instructed to identify tachistoscopically presented words as rapidly as possible. The results were similar to those reported for the functional-fixedness study: when the correct response was dominant, an increase in drive improved performance, i.e., the visual duration threshold was lowered. In contrast, increasing drive when the correct response was low in the hierarchy raised the threshold.

There is obviously still much more work, both empirical and theoretical, needed to develop the model that has been described. At this point it may be appropriate to summarize the major points of this paper.

There is not just one way to investigate problem solving. The researcher who is interested in problem solving has several different pretheoretical models from which to choose. This paper reported the results of a research program based on an S-R model in which the importance of horizontal and vertical processes were emphasized. Horizontal processes refer to the linking of successive S-R units into a behavioral chain, while vertical processes refer to the assumption that independent chains occur simultaneously. A series of experiments was reported, the implications of which supported postulating a mediational mechanism within a behavioral chain. By comparing the behavior of human subjects of different ages, as well as relating their results to lower animals, it was possible to infer that as a child matures he makes a transition from responding on the basis of a single unit S-R mechanism to a mediational one. Additional data were cited that suggest the full impact of verbal behavior on problem solving depends on developmental processes that encourage interaction between chains at different vertical levels. It was also suggested that problem solving begins in a simple concept learning task when a correct overt response from one behavioral chain occurs contiguously and adventitiously with the appropriate implicit cue from another chain. The paper was concluded by citing findings that suggested the neobehavioristic drive theory which assumes that the effect of different levels of drive depends on the position of the correct response in the habit hierarchy is applicable to a functional fixedness problem as well as a perceptual-recognition task.

If nothing else, it is hoped that the present paper demonstrates that it is possible to investigate problem solving in a systematic fashion. If more psychologists accepted this possibility and were willing to expend their research energies in the field of problem solving, progress in this area would be greater than it is today.

REFERENCES

ADAMSON, R. E. Functional-fixedness as related to problem solving. *J. exp. Psychol.*, 1952, **44**, 288–291.

BUSS, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discrimination. *J. exp. Psychol.*, 1953, **45**, 75–81.

BUSS, A. H. Reversal and nonreversal

shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

BUYTENDIJK, F. J. J. Über das Umlernen. *Arch. Néerl. Physiol.*, 1930, **15**, 283–310.

COFER, C. N. Reasoning as an associative process: III. The role of verbal responses in problem solving. *J. gen. Psychol.*, 1957, **57**, 55–58.

DUFORT, R. H., GUTTMAN, N., & KIMBLE, G. A. One trial discrimination reversal in the white rat. *J. comp. physiol. Psychol.*, 1954, **47**, 248–249.

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

DUNCKER, K. On problem solving. *Psychol. Monog.*, 1945, **58**(5, Whole No. 270).

GLUCKSBERG, S. The influence of strength of drive on functional fixedness and perceptual recognition. *J. exp. Psychol.*, 1962, in press.

GUTHRIE, E. R. *The psychology of learning.* (Rev. ed.) New York: Harper, 1952.

GUTHRIE, E. R. Association by contiguity. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 158–195.

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HARROW, M., & FRIEDMAN, G. B. Comparing reversal and nonreversal shifts in concept formation with partial reinforcement controlled. *J. exp. Psychol.*, 1958, **55**, 592–597.

HEBB, D. O. *A textbook of psychology.* Philadelphia: Saunders, 1958.

HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **37**, 511–525.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KELLEHER, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.

KENDLER, H. H. Problems in problem solving research. In, *Current trends in psychological theory: A bicentennial program.* Pittsburgh: Univer. Pittsburgh Press, 1961.

KENDLER, H. H., & D'AMATO, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

KENDLER, H. H., GLUCKSBERG, S., & KESTON, R. Perception and mediation in concept learning. *J. exp. Psychol.*, 1961, **61**, 186–191.

KENDLER, H. H., & KENDLER, T. S. Inferential behavior in preschool children. *J. exp. Psychol.*, 1956, **51**, 311–314.

KENDLER, H. H., KENDLER, T. S., PLISKOFF, S. S., & D'AMATO, M. F. Inferential behavior in children: I. The influence of reinforcement and incentive motivation. *J. exp. Psychol.*, 1958, **55**, 207–212.

KENDLER, H. H., & LACHMAN, R. Habit reversal as a function of schedule of reinforcement and drive strength. *J. exp. Psychol.*, 1958, **55**, 584–591.

KENDLER, H. H., & MAYZNER, M. S., JR. Reversal and nonreversal shifts in card sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.

KENDLER, H. H., & VINEBERG, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, **48**, 252–258.

KENDLER, T. S., & KENDLER, H. H. Reversal and nonreversal shifts in kindergarten children. *J. exp. Psychol.*, 1959, **58**, 56–60.

KENDLER, T. S., & KENDLER, H. H. Inferential behavior in children: II. The influence of order of presentation. *J. exp. Psychol.*, 1961, **61**, 442–448.

KENDLER, T. S., KENDLER, H. H., & WELLS, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

KOCH, S. *Psychology: A study of a science.* Vol. 1. *Sensory, perceptual, and physiological formulations.* New York: McGraw-Hill, 1959.

KOHLER, W. *The mentality of apes.* New York: Harcourt, Brace, 1925.

KRECHEVSKY, I. Antagonistic visual discrimination habits in the white rat. *J. comp. Psychol.*, 1932, **14**, 263–277.

LACHMAN, R. The model in theory construction. *Psychol. Rev.*, 1960, **67**, 113–129.

LURIA, A. R. The role of language in the formation of temporary connections. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957. Pp. 115–129.

MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–143.

MALTZMAN, I. Thinking: From a behavioristic point of view. *Psychol. Rev.*, 1955, **62**, 275–286.

MEYER, M. F. *The fundamental laws of human behavior.* Boston: Gorham, 1911.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

SAUGSTAD, P. An analysis of Maier's pendulum problem. *J. exp. Psychol.*, 1957, **54**, 168–179.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

THOMPSON, R. Successive reversal of a position habit in an invertebrate. *Science*, 1957, **126**, 163–164.

TOULMIN, S. *The philosophy of science.* London: Hutchinson Univer. Library, 1953.

UNDERWOOD, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59**, 209–220.

WATSON, J. B. Psychology as the behaviorist sees it. *Psychol. Rev.*, 1913, **20**, 158–177.

WODINSKY, J., & BITTERMAN, M. E. Discrimination-reversal in the fish. *Amer. J. Psychol.*, 1957, **70**, 569–575.

(Received June 24, 1960)

Psychological Review
1962, Vol. 69, No. 1, 17–33

A QUANTITATIVE THEORY OF FEELING: 1960[1]

D. R. KENSHALO AND JOHN P. NAFE
Florida State University

In recent years considerable evidence has been presented which is contrary to the theory of von Frey (1895) that qualitative differences in cutaneous sensations are dependent upon the selective action of specialized, encapsulated end organs. Much of this evidence is from histological studies of skin of various parts of the body and leads to the conclusion that encapsulated end organs cannot be found in hairy skin. Yet hairy skin exhibits sensitivity equal to or greater than that found in glabrous skin (Sinclair, Weddell, & Zander, 1952).

This paper reviews studies concerned with the manner in which nerves terminate in skin and their role in cutaneous sensitivity. The histological studies lead to the conclusion that nerve fibers supplying hairy skin terminate as bare, unencapsulated filaments in relation to three types of tissue: undifferentiated epithelial cells of the dermis and lower layers of the epidermis, hair follicles, and smooth muscle elements of the cutaneous blood vessels. Nafe (1934, 1942) has proposed that these bare filaments are essentially alike and hence respond to a common adequate stimulus, namely, movement either in relation to themselves or to the surrounding tissue. These terminals differ primarily in that they terminate in different tissues of the skin. Both tactile and thermal sensations result when tissue is moved, the former by the direct action of a mechanical stimulus on the tissue, and the latter, by movement of the thermally labile smooth muscle, especially of the arterioles.

HISTOLOGY OF SKIN RECEPTORS

Rose and Montcastle (1959) have reaffirmed the position of most textbook writers that the skin senses are mediated by specialized encapsulated endings. Commenting on the recent publications by the Oxford group (Weddell, Palmer, & Pallie, 1955), in which it is maintained that there is but a single type of nerve ending mediating cutaneous sensations, Rose and Montcastle (1959) say:

> If a crisis exists in respect to evaluating the morphology of the endings, it is a crisis of abundance and not of scarcity. One hesitates to accept as a solution to the vexing problem of the morphology of the encapsulated endings a declaration that virtually all morphological differences between them are either insignificant or due to artifacts of the technique (p. 390).

If the Oxford group were the only investigators to arrive at such conclusions, one might expect such a response as quoted above. The controversy as to whether or not encapsulated endings mediate the various qualities of cutaneous sensations has a much longer history.

Three major methods have been employed in the attack on the problem of the histology of cutaneous sensation.

---

[1] The first paper under this title was published by John P. Nafe in 1929. The present paper was presented in modified form at a symposium on Cutaneous Sensitivity at the Army Medical Research Laboratory, Fort Knox, Kentucky, February 11–13, 1960.

Supported by Contract No. DA 49-007-MD-683 between the Office of The Surgeon General, Department of Army, and Florida State University. Reproduction in part or in whole is permitted for purposes of the United States Government.

TABLE 1

SUMMARY OF HISTOLOGICAL INVESTIGATIONS RELATING SPECIALIZED ENCAPSULATED END ORGANS WITH CUTANEOUS SENSATIONS

Investigator	Skin area		Sensations tested	Encapsulated endings reported
	Glabrous	Hairy		
Type I Investigations				
Donaldson (1885)		forearm	cold	none
Goldscheider (1886)		forearm	cold, warm, touch, pain	none
Häggqvist (1913)		forearm	cold, warm	none
Dallenbach (1927)		upper arm	cold, warm	none
Pendleton (1929)		forearm	cold	none
Woolard (1935)		thigh	cold, pain, touch	none
Bazett (1941)		forearm	cold	none
Weddell (1941)		forearm	cold	Krause end bulb
Belonoschkin (1933)	nipple		cold	complicated plexuses without capsules
Type II Investigations				
von Frey (1895)	conjunctiva		cold	Krause end bulb
	cornea		cold, warm, touch, pain	none
		forearm	cold	Krause end bulb
Strughold (1925)	mouth mucosa		cold	Krause end bulb
Strughold and Karbe (1925)	conjunctiva		cold	Krause end bulb
Bazett et al. (1932)	prepuce		cold, warm, touch	Krause end bulb, Ruffini cylinders
Sinclair et al. (1952)	finger		cold, warm, touch	Meissner corpuscle, Merkel's discs, Krause and Ruffini endings
		forearm	cold, warm, touch	none
		auricle	cold, warm, touch	none
Type III Investigations				
Gilbert (1929)		thigh		none
		epigastrium		none
		chest		none
		breast		none
Gilmer and Haythorn (1941)		back		none
Hagen et al. (1953)		abdomen		none
Dastur (1955)		finger dorsum		none
Winkilmann (1955)		232 specimens from hairy skin		none
Woolard (1936)		hairy skin		none
		thigh		
Gilbert (1929)	foot sole			Meissner corpuscle Merkel's disc
Williams et al. (1929)	nipple			Meissner corpuscles Golgi-Mazzoni capsules
	prepuce			Krause end bulb Ruffini cylinders
Cathcart et al. (1948)	nipple			Meissner corpuscles 4 unnamed types
Gairns (1955), Gairns (1951)	palate			Krause end bulb (rare) unnamed smaller capsules
	gum			Meissner corpuscles Krause end bulb
Gairns (1953)	tongue			unnamed varieties Krause end bulb Meissner corpuscles

The first involves careful sensory testing and permanent marking of the points on the skin which are sensitive to thermal and/or tactile stimulation. After sensory testing and marking, the tissue is removed, stained, and examined microscopically. This is the most direct approach. The second method involves a determination of the number and kind of sensory spots found in a skin area followed by a histological examination of the tissue. The number and type of encapsulated end organs are related to the frequency and kind of sensory spots, but no attempt is made to draw a close relationship between a spot and a particular end organ as in the first method. The third only describes the general location and distribution of various kinds of encapsulated endings. The second and third methods give supportive data at best.

Investigations of the first type have been carried out at least nine times since the first report of Donaldson (1885). With the exception of the study reported by Belonoschkin (1933), these investigations have been carried out on hairy skin. It will be noted in Table 1, with one exception, that the investigators have failed to find encapsulated endings of any sort beneath the spots of stimulation. Investigations of the second type have been conducted primarily on specialized tissue, such as is found in the conjunctiva, the prepuce, and the palmer side of the finger. While there is no direct correlation between a sensory spot and the type of ending beneath it, in general, these investigators conclude that the number and type of encapsulated endings correspond rather well with the number and type of sensory spots reported in that tissue. The list of investigations of the third type included here is by no means exhaustive. The inclusion of these authors is on the basis of their interest in cutaneous sensation. It will be noted that, with one exception, when specimens of skin were removed from the hairy regions of the body, encapsulated endings were not found; whereas, when the specimens of skin were obtained from the nonhairy portions of the body, encapsulated endings were quite prevalent.

In this entire list of investigations there are three exceptions to the general conclusion that encapsulated endings do not occur in hairy skin. These are Weddell (1941), in the first type of investigation; von Frey (1895), in the second type of investigation; and Woolard (1936), in the third type of investigation. Weddell et al. (1955) now contend that the structures which he and Woolard (1936) reported as being encapsulated endings were either unusual sections of hair follicles, or artifacts which resulted from rough handling of the specimens during the fixing process. This appears to be reasonable in view of the results obtained by the other investigators. The conclusion is inescapable that encapsulated endings do not occur in hairy skin. They exist only in specialized skin areas as the conjunctiva of the eye, the mucous membrane of the mouth, the palmer surfaces of the fingers, and the palms of the hand, the soles of the feet, and the genitalia. In hairy skin nerves terminate as bare filaments. The manner of termination of these afferent nerve fibers has been investigated by Weddell, Pallie, and Palmer (1954), and the work of other investigators has been reviewed by Weddell et al. (1955). In general, afferent nerves terminate as bare filaments in relation to three types of structures found in the skin. These are: in and among the cells of the stratum granulosum of the epidermis and among the cells of the dermis; some of the largest fibers serving the skin end in relation to hair follicles, as filaments entwined about the hair shaft; and other smaller fibers terminate in relation to the smooth muscle cells of the cutaneous blood vessels.

Punctate sensitivity of the skin, as described by Blix (1884) and Goldscheider (1884) is often cited as supporting evidence for the specialized encapsulated receptor theory. Some areas or spots of skin are, indeed, more sensitive than others when near threshold intensities of stimulation are used.

However, in studies in which intensity of tactile (Guilford & Lovewell, 1936) or thermal (Jenkins, 1940) stimuli were systematically changed, the number of sensitive spots varied directly with the intensity of the stimuli. It appears that the skin is more or less continuously sensitive to tactile and thermal stimuli, exhibiting peaks and valleys of sensitivity.

CUTANEOUS SENSATION

It is no longer profitable to hypothesize encapsulated end organs which are sensitive to specific types of energy applied to the skin. Another organizing principle must be found to account for qualitative differences in sensations derived from cutaneous stimulation. The filaments in which these nerves terminate appear morphologically alike, although it cannot be maintained that they are identical (Sperry, 1950; Weiss, 1947). On the other hand, there seems to be little reason to assume that the differences are sufficiently great to account for the wide variety of sensory experiences derived from their stimulation. Since there is no reason, at present, to assume that they are different, it would be expected that the terminals of these cutaneous fibers would share a common adequate stimulus, regardless of the tissue in which they end. Evidence has been presented (Nafe & Kenshalo, 1958; Nafe & Wagoner, 1941a, 1941b) which indicates that movement is this common adequate stimulus. Differences in the qualities of cutaneous sensations are based upon differences in the properties of the tissue innervated. Tactile sensations are aroused by and during movement of dermal and epidermal tissue. Thermal sensations are also aroused by tissue movement, although in this instance, the tissue is the smooth muscle composing the walls of the cutaneous vascular system (Nafe, 1942).

The Tactile Sense

Two components exist in the application of any mechanical stimulus to the skin. These are a static component (or magnitude of displacement), and a dynamic component (or velocity and acceleration). Acceleration has not been well investigated, although other work (Hubbard, 1958) seems to indicate that it is of little significance.

The earlier theories concerned with the nature of the adequate stimulus for the tactile sense considered the static component of the mechanical stimulus to be the relevant aspect of that stimulus; thus, there is the pressure theory (Weber, 1846), the gradient theory (Meissner, 1859), and the tension theory (von Frey & Kiesow, 1899). von Frey and Kiesow formulated their theory on the basis of stimuli involving very small areas; that is, stimuli with radii of less than 0.1 millimeter. They found that tension (weight per unit radius) was related to the threshold of sensation for these very small stimuli. On the other hand, for larger stimuli, greater than 0.4-millimeter radius, they were able to relate only the rate of application of the stimulus to the threshold of sensation. Apparently, von Frey and Kiesow were so convinced of the importance of the static component of the mechanical stimulus that they ignored the rate factor involved in the stimuli of larger radii.

Nafe and Wagoner (1941a, 1941b) were the first to record the motion of a mechanical stimulus in producing a tactile sensation. Among the recorded variables of weight, area, time to cessation of the sensation, and the rate of stimulus movement at the cessation of the sensation, a relationship was established between weight and the rate of stimulator movement at the time of cessation of the tactile sensation. The rate of movement varied independently of the area at the time of cessation of sensation; hence, any aspect of the stimulus which involves this variable, e.g., pressure or tension, must be considered to be of no importance in producing and maintaining a tactile sensation. In addition, they found that depth of deformation of the tissue was a function of the applied pressure and must also be rejected as the stimulus. In other words, they were not able to relate any of the static aspects of the stimulus to the point at which cessation of the sensation occurred. As a result, they concluded that the dynamic component was the relevant aspect of the stimulus; further, that cessation of the tactile sensation was the result of failure of the dynamic component of the mechanical event to stimulate, rather than failure of the receptor to respond to the stimulus. The rate of tissue deformation had reached a subliminal value.

To test this notion, Nafe and Wagoner (1941b) lowered a weight of 17.5 grams onto the skin is 8.75-gram increments. The subject reported a tactile sensation with the application of each increment and cessation of the sensation when the rate of stimulus movement became small. Furthermore, when the second 8.75-gram weight was removed, the subject again felt a tactile sensation which disappeared at approximately the same rate of tissue movement as before. They contend that had the receptors, rather than stimulation, failed, the subject would not have felt the application of the second weight, nor would he have felt its removal.

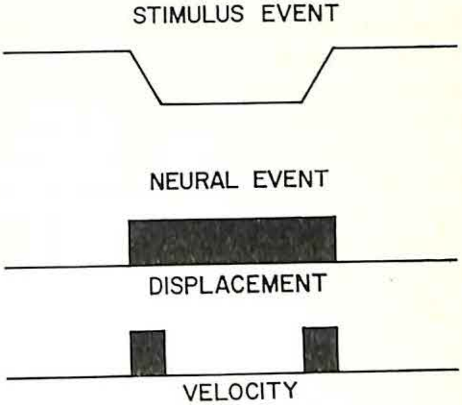


FIG. 1. Hypothetical response in a nerve fiber to a stimulus. (If the tactile receptor responds to the static component of a stimulus, action potentials should occur throughout stimulation as shown in the second line labeled "Displacement." The neural response to the dynamic component of the stimulus, in the third line, will last only during movement of the stimulus.)

In order to locate this effect in the peripheral receptor mechanism, rather than in the central nervous system, Nafe and Kenshalo (1958) repeated Nafe and Wagoner's experiments using a slightly modified version of the Nafe and Wagoner apparatus. In place of the human observer, an animal, usually a rat, was used. Instead of a verbal report by the subject, action potentials were recorded from the nerve innervating the stimulated tissue. Simultaneous recordings of the stimulus action and the neural discharge were made on photographic paper. In most of the preparations, the tongue of the rat was stimulated and recordings were made from the lingual branch of the fifth cranial nerve. Some preparations were made of the shaved leg of the rat and records were made from the femoral nerve, while in other experiments, the skin of frog was used.

Figure 1 shows the type of record which would be predicted if the receptors of the skin respond to either the static component (displacement) or the dynamic component (velocity) of the mechanical stimulus.

Figure 2 shows a record of the stim-

8-372

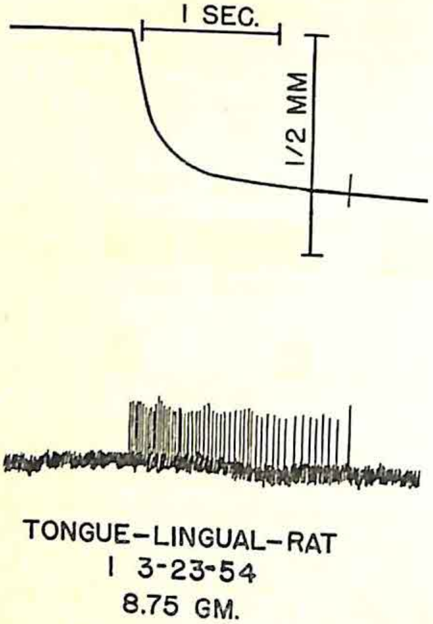


FIG. 2. Action potentials in the lingual nerve in response to weight applied to the tongue of rat. (Movement of the weight into the tissue is shown in the top line. This is comparable to only the left half of Figure 1. In other experiments the weight was later lifted. Action potentials were also obtained during its removal—modified from Nafe & Kenshalo, 1958.)

ulus resulting from an 8.75-gram weight sinking into the tissue of rat tongue, and the neural discharge which results. In this type of recording, the stimulator follows the adjustment of the tissue to the weight rather closely. It can be seen that the neural discharge which accompanied application of the stimulator stopped before the tissue had completely adjusted to the added weight. The stimulator was still moving into the tissue, in this instance at the rate of 0.019 mm/sec, when the neural discharge stopped. Records such as these establish the point that the cessation of sensation associated with a minimal rate of movement is a peripheral event.

The question still remains as to whether the cessation of discharge is a function of the stimulus effectiveness alone, or if it is in part due to adaptation by the receptor. To test this, the stimulus was hydraulically driven into the tissue in two steps of equal velocity, but the second step was of less magnitude than the first. If the cessation of discharge were associated with receptor adaptation rather than a stimulus parameter, the neural discharge associated with the second step should be less than that of the first. On the other hand, if rate of movement is the determining condition, the discharge to the second step should be equal to

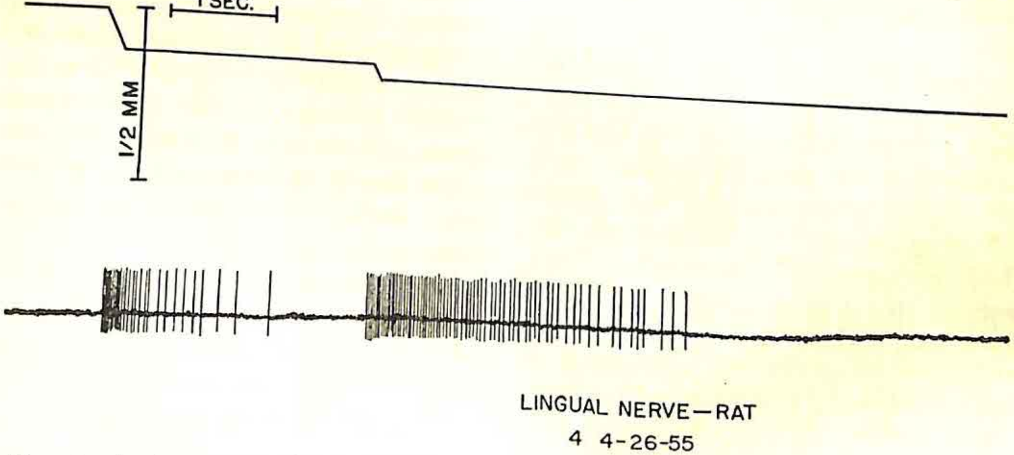


FIG. 3. Action potentials in the lingual nerve of rat when the stimulator is driven hydraulically into the tongue tissue in two steps of equal velocity. (The course of the stimulus into the tissue is shown in the top line—modified from Nafe & Kenshalo, 1958.)

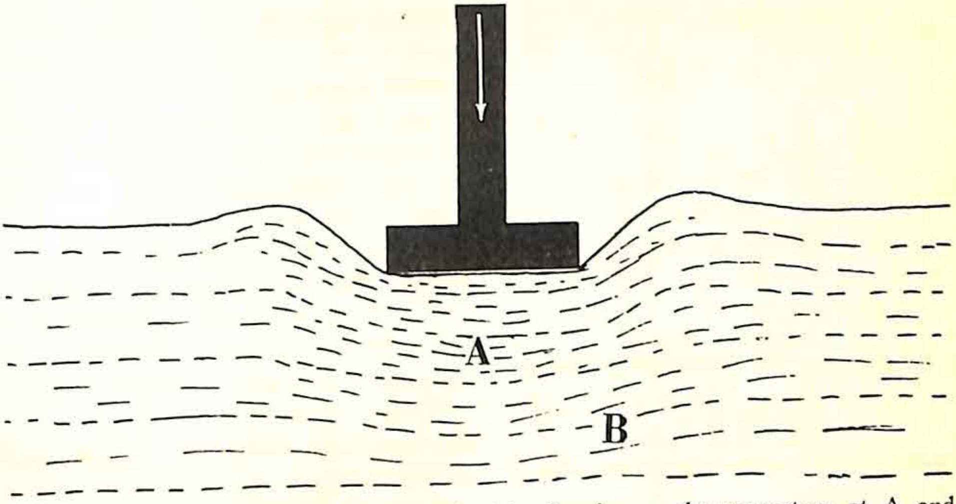


FIG. 4. Hypothetical cross section through skin showing mechanoreceptors at A and B in relation to the stimulator.

or greater than that from the first. As can be seen from Figure 3, the second hypothesis is confirmed. The longer discharge associated with the second step is explained by greater mechanical stress in the tissue following the application of the first step, than prior to its application. Thus, in Figure 4, if it is assumed that the terminals of the fiber from which the record was made are at A, it is apparent that movement of tissue at A will be greater per unit movement of the stimulator when A is under mechanical stress than when it is not. The observations of Loewenstein (1956) emphasize the importance of the mechanical factors involved in the stimulation of skin receptors. He has reported that the duration of a single unit discharge in response to a standard stimulus in frog skin is greater when the skin is placed under mechanical stress than when it is flaccid. Our own observations of a decrease in the velocity of stimulus movement with increasing weights applied over a constant area, may be explained in a similar fashion. One might think of the relation of the stimulus and receptor as analogous to a mechanical gear train with a large amount of backlash. As the gears are put under mechanical stress, by loading, e.g., the use of spring loaded antibacklash gears, the output of the gear train follows that of the input more closely. Similarly the tissue at A follows movements of the stimulator more closely.

To test these notions it seemed appropriate to use an unyielding tissue. The coupling between the stimulator and receptor would be closer and tissue readjustments to the stimulus action would be minimal compared to the rat's tongue. Frog skin is suited for such an investigation. As can be seen in Figure 5, the neural discharge occurred only during the time that the tissue was being moved. The second interesting aspect is that when the stimulator was removed from the frog's back a third series of neural impulses occurred.

Movement of tissue appears to be the adequate stimulus for nerve terminals found in a variety of locations. Nafe and Wagoner (1941b) used the skin of the web between the thumb and index finger and just above the knee. Nafe and Kenshalo (1958) used the tongue and hind leg of rat,

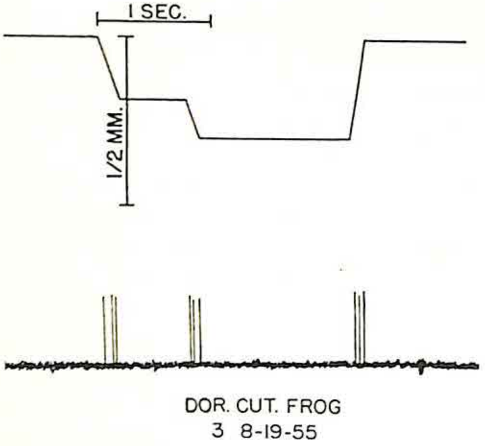


FIG. 5. Action potentials in the dorsal cutaneous nerve of frog as a result of a stimulus driven hydraulically into the skin and then removed. (The course of the stimulus into the tissue is shown in the top line—modified from Nafe & Kenshalo, 1958.)

and frog skin. There appears to be little difference in the mechanism of stimulation of the nerve terminals which exist in these various locations. Differences which have been observed may be accounted for by differences in the mechanical properties of the tissue surrounding them. It has been known for some time that nerve terminals surrounding hair follicles respond only to movement of the hair (Adrian, 1930). In addition, the works of Gray and Malcolm (1950), Gray and Matthews (1951), and Gray and Sato (1953) have demonstrated that the Pacinian corpuscle is also responsive to only the dynamic aspect of the stimulus—movement.

According to this concept, tactile adaptation, in the sense of a protracted but diminishing sensation or nerve discharge resulting from a constant stimulus, becomes meaningless. Consider the diagram of Figure 4. The rate of tissue movement or adjustment at A will be much less than the rate of movement of the stimulator, but as it continues to sink into the tissue, the rate of tissue movement will approach that of the stimulator. If it is assumed that a receptor is at A, it is clear that the stimulus to which the receptor responds is by no means constant, although the stimulus object, which the experimenter sees, is constant—it weighs 5 grams, or exerts 5 grams/cm^2 pressure. If a solid plate such as bone is placed immediately beneath the diagram, shearing forces are also produced because the tissue will be squeezed between the stimulator and the bone. The result of this added complication will be that resolution of these stresses will require much longer at A than at B. According to the older electrophysiological literature, in comparing the neural discharges from receptors at A and B, it would have been said that A was slow adapting while B was fast adapting, implying a difference in the neurological functioning of each rather than in tissue mechanics. An important parameter of the problem is overlooked if the mechanical properties of the tissue are not also investigated.

The Thermal Sense

Before considering proposals for thermal receptors, the facts concerning thermal sensation should be examined. They impose rather stringent requirements upon any theory concerned with the nature of thermal reception. The thermal sense is different from other senses. In the other senses absence of sensation is associated with near zero energy levels. Zero thermal energy is at absolute zero, approximately − 273° C; yet, psychological zero or zero thermal sensation is at skin temperature—approximately 33° C.

Stimulation of thermal receptors depends upon a temporal thermal gradient. Within limits, thermal sensations occur only when the stimulus temperature change occurs at above a minimal rate. Skin temperature is known to shift spontaneously by as much as 0.5° within the space of a few minutes (DuBois, 1941); yet during these shifts the subject is not aware of any thermal sensation. The observations of Hensel (1950) indicate that skin temperature may be shifted within the range of 29°–36° C without producing a thermal sensation, provided the change is at a rate of less than 0.007° C/sec. This appears to be the zone of complete thermal adaptation and hence specifies the limits of the shifting psychological zero. The shifting zero point accounts for the frequent observation that the thermal reception system in man is not a good absolute thermometer. While thermal sensations do not occur with large changes in skin temperature at a slow rate, threshold sensations occur at a rate of 0.001° C/sec for 3 seconds (Hardy & Oppel, 1936).

The thermal threshold is affected by the temperature of the skin (Hensel, 1950). Ebaugh and Thauer (1950) have reported that the cool threshold is increased at skin temperatures above 33° C. Kenshalo, Nafe, and Dawson (1960) confirmed their findings and, in addition, were able to show that the warm threshold decreases at skin temperatures above 34° C.

Theories which depend on a spatial thermal gradient must be rejected. Ebbecke (1917) and Bazett and McGlone (1932a) have suggested spatial gradient theories. The notions of Bernhard and Granit (1946) and von Euler (1947), that nerve itself may act as a thermal receptor, is also based on regional cooling of the nerve. Lele, Weddell, and Williams (1954) have extended this notion and suggested a Bolometer Theory which depends on warming and cooling of terminals in relation to their stem axons. Rejection of spatial gradients as a means of stimulating thermal receptors requires some justification. Hensel and Zotterman (1951c), using electrophysiological techniques, located the termination of a fiber responding to cold close to one side of cat's tongue. They found that the receptor performs equally well regardless of which side of the tongue was cooled. The older work of Bazett, McGlone, Williams, and Lufkin (1932) on the prepuce, likewise, demonstrates the ineffectiveness of the direction of the spatial thermal gradient as a means of stimulating warm receptors. Vendrik and Vos (1958) abolished a spatial thermal gradient altogether by warming the skin with a radar beam of 10-centimeter wave length, yet were unable to observe any appreciable difference in the warmth threshold from that obtained by infrared radiation. Electromagnetic wave lengths of this order are quite penetrating, according to them, and warm the deeper tissues as rapidly as the skin surface. If a spatial thermal gradient were necessary to stimulate the thermal sensing mechanism, they should have obtained a much higher thermal threshold using the radar beam than by infrared stimulation. The latter raises the skin surface temperature and depends upon conduction of the thermal energy to the deeper tissues.

These experiments demonstrate the ineffectiveness of spatial thermal gradients in depth as a stimulus for the thermal sensing mechanism. It is still possible that gradients across the surface of the skin may be of some importance, although it seems unlikely. Experiments such as those by Hardy and Oppel (1936) in which large areas of skin, e.g., the torso or torso plus face, were warmed uniformly by radiant energy should have produced thermal sensations only at the edges of the stimulus. Their subjects made no such observation. Similarly, the thermal sensation aroused by laying a cold coin on the skin should occur only at the margin, but this is not consistent with common observation.

Two types of receptor mechanisms have been proposed for mediating thermal sensations. One assumes that cutaneous nerve terminals are differentially sensitive, some responding only to mechanical energy, while others respond to thermal energy. The other maintains that endings of the skin respond to a common adequate stimulus—movement. Thermal reception would then be by means of tissue movement resulting from thermal stimulation. It is well known that smooth muscle is thermally responsive, contracting when cooled and relaxing when warmed. Such movement induced in smooth muscle of the cutaneous vascular system would produce impulses in the nerve fibers terminating among these smooth muscle elements.

The notion that nerve terminals respond differentially to mechanical and thermal energy is supported by evidence from Hensel and Zotterman (1951b). They found that stem axons which show activity in response to the application of mechanical stimuli to their terminals do not respond well to thermal stimuli and vice versa. Their findings indicate that mechanically sensitive nerve terminals have stem axons of larger diameter (larger than 8 micra); whereas, those responding to thermal stimuli are smaller (1.5–5 micra). They interpret their findings as indicating that nerve terminals are inherently different in their responsiveness to thermal and mechanical energy. However, such differences could occur as a result of differences in the mechanical and thermal properties of the tissue in which the nerve terminals end.

Several mechanisms have been suggested to account for the apparent differential sensitivity of nerve terminals. Bishop (1946) has related it to characteristics associated with the size of the stem fiber involved. Others have suggested that the nerve terminals differ in their sensitivity to chemical changes. Bazett and McGlone (1932b) proposed that thermal sensations might be the result of alterations in blood acidity produced by warming and cooling. Consistent with this, Dodt (1956) has demonstrated a differential effect of CO_2 tension on the response of nerves stimulated thermally. In his Concentration Theory, Jenkins (1941) has also suggested a chemical intermediary. Bare nerve endings, according to Jenkins, would discriminate between a catabolic and an anabolic phase of some chemical reaction brought about by a change in temperature. Presumably bare nerve endings responding to mechanical stimuli would be insensitive to these changes.

The suggestion of Hensel and Zotterman (1951b) on the mechanism of the stimulation of cold receptors is reminiscent of the anabolic-catabolic hypothesis. Because of the rather complex shifts in the frequencies of impulses which they found in fibers responding to a temperature reduction, they suggest that there are two temperature dependent processes, probably chemical, involved in the stimulation of cold fibers (Zotterman, 1959). The impulse frequency resulting from cold stimulation is a function of a difference in the activity of these two processes. Although they do not speculate on the neural response to warming, one must presume at least one, and probably two, additional processes are necessary. Dodt (1953) has shown that the responses of mammalian A fibers to warming and cooling are functions of their diameter. The differences do not appear great enough, however, to aid in understanding Hensel and Zotterman's suggestion of a chemical process controlling the responsiveness of certain fibers to thermal stimuli.

Hypotheses which attempt to explain differences between thermal and mechanical sensitivity on the basis of nerve terminals of different sensitivity must assume two different sets of cutaneous nerve terminals, one set responding to a mechanical form of energy, the other to thermal energy. If the difference between the two sets of terminals is a chemical one, the chemical reaction should be one which is in equilibrium at the mid-range of the thermal energy continuum, which may have its thermal equilibrium point shifted by 6 or 7 degrees if done sufficiently slowly, yet which will become unstable when changed rapidly by as little as 0.003° C (Hardy & Oppel, 1936). To further complicate the theory, Hensel and Zotterman (1951b) find that of the terminals responding to thermal energy, some are activated only by an elevation, while others respond to a reduction in the thermal energy. Thermal fibers must now be endowed with even greater selectivity, either to discriminate between two different chemical reactions, one for warming and one for cooling, or between the directions in which a reversible chemical reaction is driven. In addition, the response of terminals to cooling is to be explained by the interacting of two temperature dependent processes, according to Zotterman (1959).

An alternative, involving fewer assumptions, is that nerve fibers and their terminals are largely alike; further that the differentiation necessary to account for thermal and mechanical responsiveness occurs as a direct result of the tissue in which the fibers terminate. It has been demonstrated that nerve terminals, similar to those terminating among the cells of the epidermis, end among the smooth muscle fibers of the cutaneous vascular network (Weddel et al., 1954; Woolard, 1926). Smooth muscle is known to be thermally responsive, relaxing upon warming and constricting when cooled. Furthermore, Nafe and Kenshalo (1958) have demonstrated that movement of the nerve terminal either upon itself or in relation to its surrounding tissue best fits the facts concerning mechanical reception.

Goldscheider (1886) was probably the first to suggest a relationship between the cutaneous vascular system and thermal sensation. He noted in his biopsies of temperature sensitive spots the "striking and immediate proximity" of blood vessels to cold spots which he has previously mapped. Elaborating on this, Nafe (1934) pointed out the striking relationship between changes in the smooth muscle response to temperature and the subjective reports of stimulation by the same temperatures. The figures which Nafe (1934) cited applied to smooth muscle of the uterus and while they apply to smooth muscle in general, recent evidence indicates that the smooth muscle of the cutaneous vascular network, especially in the venules,

> show an extraordinarily prominent vasomotor activity in contrast to other tissues, such as mesenteric structures, the wall of the intestinal tract, and skeletal muscle, where venous vasomotion is minimal and seemingly passive (Zweifach, 1959, p. 69).

Furthermore, the venular tributaries in the skin are ten- to twenty-fold more sensitive to a fall in temperature of 1–2° C than other tissues (Zweifach, 1959). While the Vascular Theory is usually thought of as involving only the arterioles the responsiveness of venular smooth muscle and that of the hair erector muscles should not be overlooked.

The Vascular Theory accounts for the requirements of a thermal theory as outlined at the beginning of this section. It does not depend upon spatial thermal gradients to achieve stimulation and is uniquely suited for explanations involving temporal thermal gradients. Reference was made earlier in the paper to the fact that in order to stimulate nerve terminals located among the cells of the epidermis and hair follicles, minimal rates of tissue movement were necessary. On the assumption that nerve terminals are essentially alike, the same conditions of stimulation should apply to the nerve terminals ending in relation to the fibers of the smooth muscle of the cutaneous vascular system. The shifting psychological zero with no attendant thermal sensation is accounted for in terms of the constriction or dilation of the cutaneous vessels occurring slowly enough to be below the threshold rate of movement necessary to excite the nerve terminals in the vessel walls. Complete thermal adaptation represents the adjustment of vascular relaxation or constriction from one level to another; and the range through which Hensel (1950) has reported that complete thermal adaptation is possible, represents the range through which the smooth muscle of the cutaneous vessels can adjust and adopt a relatively stable end point. Outside of this range vascular smooth muscle would be unable to achieve a relatively stable adjustment to the temperature.

Changes in the size of the thermal threshold as a function of the skin temperature should be expected since the latter would govern the degree of relaxation or constriction of the smooth muscle. Predictions of the direction in which the thresholds would go are difficult to make because of ignorance of the mechanical and physiological characteristics of afferent innervation of smooth muscle.

In addition to fitting these requirements, the theory also handles facts which have been somewhat perplexing to other theories. Thermal sensations occur under some circumstances which do not include a change in the environmental temperature. It has been noted, especially in the earlier work on the thermal sense, that mechanical stimulation of the skin may produce sensations of warm or cool (Nafe, 1934). Smooth muscle is known, upon occasion, to relax or contract when prodded, depending upon its state of contraction at the moment (Evans, 1956). Thermal sensations resulting from mechanical stimulation can be explained in this manner if one assumes that they are mediated by smooth muscle of the cutaneous vascular system. Warm sensations accompanying a blush occur as a result of chemically mediated vasodilation since there is apparently little if any reflex vasomotor control in the "blush area" (Burton, 1959). The chill accompanying scratching a fingernail across a blackboard is the result of a reflexly aroused general vasoconstriction.

It is not necessary to postulate different types of free nerve endings, each responsive to one type of energy; nor is it necessary to conceive of two different systems, one for warm and one for cool. Two possible explanations exist. One depends upon neural patterns while the other depends upon different fibers being stimulated, based upon the way in which they terminate among the smooth muscle elements. In the first instance neural patterns aroused by constriction could conceivably differ from those aroused by relaxation. Patterns in other systems result in different sensations. For example, the peristaltic waves in the stomach passing from the cardiac to the pyloric valves are associated with hunger pangs, while those same contractions passing in the opposite direction result in the entirely different sensation of nausea. The direction of travel is simply a matter of timing. In addition to central factors involved in thirst, the patterns aroused in the membranes of the throat when they are dry, in comparison to when they are wet, contribute markedly to this sensation. The urge to micturate and to defecate are likewise results of patterning of neural discharges.

On the other hand, Hensel and Zotterman (1951c), using electrophysiological methods on the cat's tongue, have described two apparently different sets of nerve fibers, one responding to warm and the other to cool stimuli. In striated muscle, such systems occur; one responding when the muscle is contracted and the other when it is stretched. In the first instance the receptor elements are connected in series with the muscle fibers; whereas, in the latter case they are connected in parallel with the muscle fiber. Whether the same type of endings exist in smooth muscle has not been investigated, although Fischer (1944) has described some afferent terminals in smooth muscle as "spray-like," and others as "spindle-like."

Hensel and Zotterman (1951b) have described fibers responding to thermal stimulation as smaller (1.5–5 micra) than those responding to tactile stimulation (8 micra and larger). Weddell and Pallie (1954) have shown that many of the fibers innervating the cutaneous blood vessels are of dorsal root origin. In addition, these fibers are among the smallest found in the skin, being, for the most part, less than 3 micra in diameter. The fibers ending in cutaneous vessels, then, correspond with the fiber spectrum which Hensel and Zotterman have described as responsible for thermal sensitivity.

One of the most frequently cited reasons for considering warm and cool to be mediated by different receptor mechanisms is the long reaction time of an observer to warm compared with cool stimuli. Bazett et al. (1932) and Hensel and Zotterman (1951a) have used this difference in latency to calculate the depth of the receptors. These calculations are based on the assumption that the latency of response of the receptors are the same but occur at different depths. An alternate explanation, in terms of the Vascular Theory, is equally possible. Constriction is an active process requiring little time to accomplish. On the other hand, dilation is a passive response and requires a much longer time to reach a threshold rate of movement.

Paradoxical cold provides little difficulty for the Vascular Theory. Paradoxical cold is a term ascribed to the situation in which a cold sensitive area is stimulated by means of a high temperature and produces a cold sensation. At high temperatures smooth muscle has constricting elements within a generally relaxing system (Evans, 1956). It might be expected that an area normally responding to a cool stimulus would also produce a cool sensation when stimulated by a high temperature. It is interesting to note that Zotterman (1953) reports fibers which normally respond to cooling, after a period of no response to temperature above 35° C. suddenly start to respond when stimulated by temperatures of 45° C. and higher.

Sensations of heat are the result of a mixed sensation of warm and paradoxical cold according to Alrutz

(1900). It is known (Evans, 1956) that when smooth muscle is warmed to above approximately 45° C. elements of the muscle begin to constrict while the main body of the muscle continues to relax. According to the Vascular Theory hot sensations result from the neural activity brought about by the constricting elements in the generally relaxing smooth muscle of the cutaneous arterioles.

The question arises as to how the sensorium distinguishes between tissue movements induced by mechanical stimuli and those induced by thermal stimulation. Evidence has already been presented to indicate that there is strong likelihood that different afferent nerve fibers respond to mechanical stimulation than to thermal. Terminals ending in the cutaneous vascular system, because of their smaller size and probably also because of the mechanical properties of the tissue of blood vessels, do not respond as readily to mechanical stimuli, as those which terminate among the cells of the epidermis, dermis, and in relation to hair follicles. There is no contention that nerve or its terminals may not be stimulated directly by a temperature change, but it has never been demonstrated that their temperature sensitivity is sufficient to account for thermal sensations in the skin. The coding system is based on different fibers being activated by different types of energy as the result of the mechanical and thermal properties of the tissue in which their terminals end. Further differentiation is accomplished in the spinal cord. Descriptions of the neuroanatomy of the spinal cord (Ranson & Clark, 1959) maintain that fibers serving touch travel in the fasciculi gracilis and cuneatus as well as in the ventral spinothalamic tract, while those allied with thermal sensation and pain ascend the cord predominantly in the lateral spinothalamic tract, although there is probably considerable intermingling of fibers among the various tracts (Rose & Montcastle, 1959). Fibers from these tracts terminate in different brain stem nuclei but their connections are beyond the scope of this paper.

Aside from the fact that the theory can account for a large number of the facts concerning tactile and thermal sensitivity it has the added advantages of simplicity and parsimony. It can serve as a convenient model within which to organize additional facts as well as bringing into sharp relief gaps in present knowledge. Furthermore, such a model allows predictions and suggests direct experiments which should be done.

SUMMARY

Qualitative theories requiring specialized encapsulated end organs or specialized peripheral nerve fibers to account for the various qualities (pressure, warm, and cold) of cutaneous sensation fail to present direct evidence that such specialization exists in peripheral nerves or in their terminations. The Quantitative Theory, discussed here, maintains that qualities of cutaneous sensation are, in part, a function of the mechanical and thermal properties of the tissue in which the cutaneous afferent nerves terminate and variations in the temperal and spatial patterns of the afferent nerve discharge. The latter variations are manifest in the frequency and duration of the impulses, and the area over which they arise as well as the relative number of fibers actuated within the area. The Quantitative Theory consists of four postulates.

1. Nerves which end freely, without encapsulation of any sort, are the primary neurological elements present in hairy skin capable of mediating

sensations derived from normal stimulation of the skin. Nerves ending freely in the skin terminate in relation to different tissues.

a. Some fibers branch and terminate among the dermal and epidermal cells of the stratum granulosum.

b. Others, and probably some branches of those terminating as in (a) form a complex net about the hair follicles.

c. Smaller fibers branch and terminate among the smooth muscle elements in the walls of the cutaneous vascular system and perhaps the hair erector muscles.

2. The terminals of nerve fibers supplying the skin are functionally similar, differing primarily by the tissue in which they terminate. Being so, they share a common adequate stimulus which is movement of the terminals either upon themselves or in relation to their immediately surrounding tissue.

3. Smooth muscle responds to thermal changes, in general, relaxing when warmed but constricting when cooled.

4. The thermal and mechanical properties of the tissue in which the terminals end determine which fibers will respond to mechanical stimulation and which will be thermally sensitive. Distinction between thermal and tactile sensations depends upon the central connections which the fibers make. Qualitative differences in sensations derived from various mechanical stimuli are the result of differing patterns of neural excitation. Thus stroking a smooth surface results in different movement patterns in the tissue and hence different patterns of action potentials from stroking a rough surface. Distinction of warm and cool sensations results from either different patterns of action potentials caused by vasoconstriction as compared to vasodilation, or different fiber terminals responding to vasoconstriction than to vasodilation.

REFERENCES

Adrian, E. D. The effects of injury on mammalian nerve fibers. *Proc. Roy. Soc. Lond., Ser. B*, 1930, **106**, 596–618.

Alrutz, S. Studien auf dem Gebiete den Temperatursinne. *Skand. Arch. Physiol.*, 1900, **10**, 340.

Bazett, H. C. Temperature sense in man. In, *Temperature: Its measurement and control in science and industry.* New York: Reinhold, 1941. Pp. 489–501.

Bazett, H. C., & McGlone, B. Studies in sensation: II. The mode of stimulation of cutaneous sensations of cold and warmth. *Arch. Neurol. Psychiat.*, 1932, **27**, 1031–1069. (a)

Bazett, H., & McGlone, B. Studies in sensation: III. Chemical factor in the stimulation of end-organ giving temperature sensations. *Arch. Neurol. Psychiat.*, 1932, **28**, 71–91. (b)

Bazett, H. C., McGlone, B., Williams, R. G., & Lufkin, H. M. Sensation: I. Depth, distribution and probable identification in the prepuce of sensory end-organs concerned in sensations of temperature and touch; thermometric conductivity. *Arch. Neurol. Psychiat.*, 1932, **27**, 489–517.

Belonoschkin, B. Ueber die Kaltreceptoren der Haut. *Z. Biol.*, 1933, **93**, 487–489.

Bernhard, C. G., & Granit, R. Nerve as model temperature end-organ. *J. gen. Physiol.*, 1946, **29**, 257–265.

Bishop, G. Neural mechanisms of cutaneous sense. *Physiol. Rev.*, 1946, **26**, 77–102.

Blix, M. Experimentelle Beiträge zur Losung der Frage über die specifische Energie der Hautnerven. *Z. Biol.*, 1884, **20**, 141–156.

Burton, A. C. Physiology of cutaneous circulation, thermoregulatory functions. In S. Rothman (Ed.), *The human integument.* Washington, D. C.: American Association for the Advancement Science, 1959.

Cathcart, E. P., Gairns, F. W., & Garven, H. S. D. The innervation of the human quiescent nipple, with notes on pigmentation, erection, and hyperneury. *Trans. Roy Soc. Edinb.*, 1948, **61**, 699.

Dallenbach, K. The temperature spots and end-organs. *Amer. J. Psychol.*, 1927, **39**, 402–427.

DASTUR, D. K. Cutaneous nerves in leprosy: The relationship between histopathology and cutaneous sensibility. *Brain,* 1955, **78**, 615–633.

DODT, E. Differential thermosensitivity of mammalian A-fibres. *Acta physiol. Scand.,* 1953, **29**, 91–108.

DODT, E. Die Aktivität der Thermoreceptoren bie nicht-thermischen Reizen bekannter thermoregulatorischer Wirkung. *Pflüg. Arch. ges. Physiol.,* 1956, **263**, 188–200.

DONALDSON, H. H. Research on the temperature sense. *Mind,* 1885, **10**, 399–416.

DUBOIS, E. F. The temperature of the human body in health and disease. In, *Temperature: Its measurement and control in science and industry.* New York: Reinhold, 1941.

EBAUGH, F., & THAUER, R. Influence of various environmental temperatures on the cold and warmth thresholds. *J. appl. Physiol.,* 1950, **3**, 173–182.

EBBECKE, U. Über die Temperaturemfindungen in ihrer Abhängigkeit von der Hautdurchbluting und von den Reflexzentren. *Pflüg. Arch. ges. Physiol.,* 1917, **169**, 395–462.

EVANS, C. L. *Principles of human physiology.* (12th ed.) Philadelphia: Lea & Febiger, 1956.

FISCHER, E. Vertebrate smooth muscle. *Physiol. Rev.,* 1944, **24**, 467–490.

GAIRNS, F. W. The sensory nerve endings of the human palate. *Quart. J. exp. Physiol.,* 1955, **40**, 40–48.

GAIRNS, S. W. Nerve endings in the human gum and hard palate. *J. Physiol.,* 1951, **115**, 70.

GAIRNS, S. W. Sensory nerve endings other than taste buds in the human tongue. *J. Physiol.,* 1953, **121**, 33.

GILBERT, R. W. Dermal sensitivity and the differentiated nerve terminations of the human skin. *J. gen. Psychol.,* 1929, **2**, 445–459.

GILMER, B. VON H., & HAYTHORN, S. R. Cutaneous pressure-vibration spots and their underlying tissues. *Arch. Neurol. Psychiat.,* 1941, **46**, 621–628.

GOLDSCHEIDER, A. Die specifische Energie der Temperaturnerve. *Mschr. prakt. Dermatol.,* 1884, **3**, 53–76.

GOLDSCHEIDER, A. Histologische Untersuchungen über die Endigungen der Hautsinnesnerven beim Menschen. *Arch. Anat. Physiol., Physiol. Section,* 1886, 191–227.

GRAY, J. A. B., & MALCOLM, J. L. Initiation of nerve impulses by mesenteric pacinian corpuscles. *Proc. Roy. Soc. Lond., Ser. B,* 1950, **137**, 96–114.

GRAY, J. A. B., & MATTHEWS, P. B. C. Response of pacinian corpuscles in the cat's toe. *J. Physiol.,* 1951, **113**, 475–482.

GRAY, J. A. B., & SATO, M. Properties of the receptor potential in pacinian corpuscles. *J. Physiol.,* 1953, **122**, 610–636.

GUILFORD, J. P., & LOVEWELL, E. M. The touch spots and the intensity of the stimulus. *J. gen. Psychol.,* 1936, **15**, 149–159.

HAGEN, E., KNOCHE, H., SINCLAIR, D., & WEDDELL, G. The role of specialized nerve terminals in cutaneous sensibility. *Proc. Roy. Soc. Lond., Ser. B,* 1953, **141**, 279–287.

HÄGGQVIST, G. Histophysiologische Studien über die Temperatursinne der Haut des Menschen. *Anat. Anz.,* 1913, **45**, 46–63.

HARDY, J., & OPPEL, T. The thermal response of skin to radiation. *Physics,* 1936, **7**, 466–479.

HENSEL, H. Temperaturempfindung und intracutane Wärmebewegung. *Pflüg. Arch. ges. Physiol.,* 1950, **252**, 165.

HENSEL, H., & ZOTTERMAN, Y. Action potentials of cold fibres and intracutaneous temperature gradient. *J. Neurophysiol.,* 1951, **14**, 377–385. (a)

HENSEL, H., & ZOTTERMAN, Y. Quantitative Beziehungen zwischen der Entladung einzelner Kaltfasern der Temperatur. *Acta physiol. Scand.,* 1951, **23**, 291. (b)

HENSEL, H., & ZOTTERMAN, Y. Response of mechanoreceptors to thermal stimulation. *J. Physiol.,* 1951, **115**, 16–24. (c)

HUBBARD, S. J. A study of rapid mechanical events in a mechanoreceptor. *J. Physiol.,* 1958, **141**, 198–218.

JENKINS, W. Studies in thermal sensitivity: XIV. Further evidence on the effects of stimulus temperature. *J. exp. Psychol.,* 1940, **27**, 76.

JENKINS, W. Studies in thermal sensitivity: XVII. The topographical and functional relations of warm and cold. *J. exp. Psychol.,* 1941, **29**, 511–516.

KENSHALO, D. R., NAFE, J. P., & DAWSON, W. W. A new method for the investigation of thermal sensitivity. *J. Psychol.,* 1960, **49**, 29–41.

LELE, P. P., WEDDELL, G., & WILLIAMS, C. M. The relationship between heat transfer, skin temperature and cutaneous sensibility. *J. Physiol.,* 1954, **126**, 206–234.

LOEWENSTEIN, W. R. Modulation of cutaneous mechanoreceptors by sympathetic stimulation. *J. Physiol.,* 1956, **132**, 40–60.

MEISSNER, G. Untersuchunger über den Tastsinn. *Z. rat. Med.*, 1859, **7**, 92–118.

NAFE, J. P. A quantitative theory of feeling. *J. gen. Psychol.*, 1929, **2**, 199–211.

NAFE, J. P. The pressure, pain, and temperature senses. In C. A. Murchison (Ed.), *A handbook of general experimental psychology*. Worcester, Mass.: Clark Univer. Press, 1934. Pp. 1037–1087.

NAFE, J. P. Toward the quantification of psychology. *Psychol. Rev.*, 1942, **49**, 1–18.

NAFE, J. P., & KENSHALO, D. R. Stimulation and neural response. *Amer. J. Psychol.*, 1958, **71**, 199–208.

NAFE, J. P., & WAGONER, K. S. The nature of pressure adaptation. *J. gen. Psychol.*, 1941, **25**, 323–351. (a)

NAFE, J. P., & WAGONER, K. S. The nature of sensory adaptation. *J. gen. Psychol.*, 1941, **25**, 295–321. (b)

PENDLETON, C. R. The cold receptor. *Amer. J. Psychol.*, 1929, **40**, 353–371.

RANSON, S. W., & CLARKE, S. H. *The anatomy of the nervous system.* (10th ed.) Philadelphia: Saunders, 1959.

ROSE, J. E., & MONTCASTLE, V. B. Touch and kinesthesis. In J. Field (Ed.), *Handbook of physiology*. Vol. 1. Washington, D. C.: American Physiological Society, 1959.

SINCLAIR, D., WEDDELL, G., & ZANDER, E. The relationship of cutaneous sensibility to neurohistology in the human pinna. *J. Anat.*, 1952, **86**, 402–411.

SPERRY, R. W. Neuronal specificity. In P. Weiss (Ed.), *Genetic neurology*. Chicago: Univer. Chicago Press, 1950.

STRUGHOLD, H. Die Topograhie des Kältesinnes in der Mundhöhle. *Z. Biol.*, 1925, **83**, 515–534.

STRUGHOLD, H., & KARBE, M. Die Topographie des Kältesinnes auf Cornea and Conjunctiva. *Z. Biol.*, 1925, **83**, 189–200.

VENDRIK, A. J. H., & VOS, J. J. Comparison of the stimulation of the warmth sense organ by microwave and infrared. *J. appl. Physiol.*, 1958, **13**, 435–444.

VON EULER, C. Selective responses to thermal stimulation of mammalian nerves. *Acta physiol. Scand.*, 1947, **14**, Suppl. 45.

VON FREY, M. Beiträge zur Sinnesphysiologie der Haut. *Ber. König. Säch. ges. Wiss., Math-Phys, Leipzig.*, 1895, **47**, 166.

VON FREY, M., & KIESOW, F. Ueber die Function der Tastkörperchen. *Z. Psychol.*, 1899, **20**, 126–163.

WEBER, E. Der Tastsinn und das Gemeingefuhl. *Wagner's Handworterb. Physiol.*, 1846, **3**, 481.

WEDDELL, G. The pattern of cutaneous innervation in relation to cutaneous sensibility. *J. Anat.*, 1941, **75**, 346–367.

WEDDELL, G., & PALLIE, W. Observations of the neurohistology of cutaneous blood vessels. In G. E. Wolstenholme & J. S. Freeman (Eds.), *Peripheral circulation in man.* (A Ciba Foundation Symposium) London: Churchill, 1954.

WEDDELL, G., PALLIE, W., & PALMER, E. The morphology of peripheral nerve terminations in the skin. *Quart. J. micro. Sci.*, 1954, **95**, 483–501.

WEDDELL, G., PALMER, E., & PALLIE, W. Nerve endings in mammalian skin. *Biol. Rev.*, 1955, **30**, 159–195.

WEISS, P. A. The problem of specificity in growth and development. *Yale J. Biol. Med.*, 1947, **19**, 235–278.

WILLIAMS, R., MCGLONE, B., & BAZETT, H. A comparison of the histology and physiology of the sensory end-organs of the prepuce. *Amer. J. med. Sci.*, 1929, **178**, 887–888.

WINKILMANN, R. K. Some sensory nerve endings in the skin. *Arch. Dermatol. Syphilol.*, 1955, **71**, 373–378.

WOOLARD, H. H. The innervation of blood vessels. *Heart*, 1926, **13**, 319–336.

WOOLARD, H. H. Observations on the termination of cutaneous nerves. *Brain*, 1935, **58**, 352–367.

WOOLARD, H. H. Intra-epidermal nerve endings. *J. Anat., Lond.*, 1936, **71**, 54–60.

ZOTTERMAN, Y. Special senses: Thermal receptors. *Annu. Rev. Physiol.*, 1953, **15**, 357–372.

ZOTTERMAN, Y. Thermal sensation. In J. Field (Ed.), *Handbook of physiology*. Vol. 1. Washington, D. C.: American Physiological Society, 1959. Pp. 431–458.

ZWEIFACH, B. W. Structural aspects and hemodynamics of microcirculation in the skin. In S. Rothman (Ed.), *The human integument.* Washington, D. C.: American Association for the Advancement of Science, 1959.

(Received July 21, 1960)

Psychological Review
1962, Vol. 69, No. 1, 34–53

AN ASSOCIATION MODEL FOR RESPONSE AND TRAINING VARIABLES IN PAIRED-ASSOCIATE LEARNING[1]

GORDON H. BOWER

Stanford University

This paper reports an attempt to apply a simple association learning model to the analysis of the influence of response and training variables on paired-associate learning. The issues under investigation are old ones but have not been resolved satisfactorily in the past. With the aid of the elementary learning model, the problems are posed clearly and, the data willing, adequately resolved by the use of a few simple and intuitively compelling assumptions about learning.

The first problem that led to this investigation concerns the relationship betwen the number of response alternatives (*N*) and error rate in paired-associate learning. Experimental results (Noble, 1955; Riley, 1952) are in agreement in showing that the number of errors subjects make before reaching some criterion of learning is greater the larger the number of response alternatives. There is little agreement, however, as to the interpretation of this fact.

Two possible factors could be involved. First, the effectiveness of a reinforced trial in increasing performance (i.e., the learning rate constant) may be influenced by *N;* and second, *N* may influence the probability of being correct by sheer guessing on items that are yet unlearned. It is a reasonably safe assumption that *N* has the second effect on chance guessing. Previous data are unclear on whether *N* also influences the first factor, the effectiveness of a reinforced trial. This is not an easy question to answer since it is difficult to separate the effects of these two factors in the data. Guessing occurs only on unlearned items but there is no way to tell by direct observation just how many items have been learned and how many have been guessed correctly on any given trial. Moreover, there appears to be little hope that more refined or ingenious experimental procedures will enable us to unconfound these two factors so that we may crucially test the hypothesis that *N* affects the learning rate.

This is the type of situation in which a theoretical model of learning can make a strategic contribution. Indeed, without the aid of some formal model of the learning process, the question of the effect of *N* can neither be posed clearly nor answered clearly. With the aid of a theory one can make suitable allowance for the guessing factor and thus make an assessment of the contribution of learning and guessing at every stage of the experiment.

The model to be presented has been formally treated in more detail in a previous paper (Bower, 1961). Here the theory will be presented informally and only those implications relevant to the present discussion will be introduced. For a more formal statement of the axioms and theorems, the reader may consult the prior report. The basic notion of the model is the assumption that each stimulus item and its correct response become associated on an all-or-none basis. Considering a single item, it can be in either of

---

[1] This investigation was supported in part by a grant, M-3849, from the National Institute of Health, United States Public Health Service.

two states at the beginning of each trial: conditioned to its correct response or not conditioned. If the item is conditioned at the beginning of a trial, then the correct response occurs. If the item is not conditioned, then the probability of the correct response depends somewhat upon the experimental procedure. In experiments by the writer, the subjects were told the N responses (Integers 1, 2, . . . N) available to them and were told to respond on every trial regardless of whether they knew the correct answer. If the N numbers occur equally often as the to-be-learned responses to the items, then the probability that the subject will guess correctly on an unlearned item is $1/N$ on the simplest assumptions;[2] correspondingly, his probability of guessing incorrectly is $1 - 1/N$. The following discussion of the model is oriented specifically towards such an experimental procedure.

With the theory formulated in this way, one cannot uniquely specify the subject's state of conditioning on a given item from the knowledge that he made a correct response, since this correct response may have come about by guessing on an unlearned item. Thus, the theory does not imply that the first correct response will be followed by correct responses on all subsequent trials. However, if the subject makes an error, then we can make a determinate inference that the item was not conditioned at the beginning of the trial.

We have introduced the notions of the two states of conditioning and the probability of the correct response given the item's state of conditioning at the beginning of each trial. We assume that each item in initially unconditioned and that the effect of successive reinforced trials is to provide repeated opportunities for the item to become conditioned. The single parameter of the theory is the learning rate constant (c) which represents the probability that an unconditioned item becomes conditioned as the result of a single reinforced trial. The probability that a reinforced trial fails to condition the correct response to an unlearned item is $1 - c$. The probability that an item is still not conditioned after n reinforced trials is $(1 - c)^n$. If c is larger than zero, then this probability approaches zero as n becomes large; that is, given a large number of reinforced trials, it it certain that the item will become conditioned on some one of these reinforced trials. When the item becomes conditioned, the probability of a correct response jumps from $1/N$ up to 1.

From the considerations above we

---

[2] The arguments to be developed apply only to those conditions in which the response alternatives are immediately available to the subject and he is permitted time to give a relevant response on each trial before the terminal event (e.g., reinforcement) occurs. This is the procedure most familiar to those who run animal subjects in choice situations. A more frequent procedure in human learning experiments has been to give the subject a fixed time interval in which he may respond, and the terminal event (e.g., correct response in paired associates, next item in a serial list) follows after that time interval regardless of whether the subject makes a relevant response. To remark on the sociology of experimenters, there seems to be an empirical correlation between the use of experimenter vs. subject-controlled exposure time and whether the relevant responses per se (as well as the S-R associations) must be learned in the experimental situation: when the responses must be learned, experimenters usually control exposure time; when the response alternatives are immediately available, either via instruction or construction, the subject is allowed to control exposure time. Clearly, however, there is no necessary entailment between these procedural aspects, and one may expect the empty cells of the 2×2 contingency table to be filled by future experimentation.

may obtain an expression for q_n, the probability that the subject responds incorrectly to a given item on Trial n of the experiment. To obtain q_n we note that the probability that the item has failed to be conditioned during the preceding $n-1$ trials is $(1-c)^{n-1}$; and if the item is not conditioned then the probability of guessing incorrectly is $1-1/N$. Hence, the probability of an error on Trial n is given by the product of these two factors:

$$q_n = (1-1/N)(1-c)^{n-1} \quad [1]$$

Using this elementary association model, the first question of this investigation can be restated in a clear manner: Is c independent of N? If c is independent of N, then, given an estimate of c obtained from one group, we should be able to predict in advance the performance of other groups trained with differing numbers of response alternatives by simply adjusting the N factor in Equation 1. To the extent that these free predictions are accurate, we would have good evidence for the assertion that N affects guessing probabilities but has no influence on learning rate per se.

The second question prompting this investigation concerns the relation between several different training conditions and performance in paired-associate learning. The two major conditions of training studied to date are the correction and noncorrection conditions. The more frequently used condition is the correction procedure in which subject is informed of the correct response on every trial regardless of whether he responds correctly. In noncorrection training, the subject is told right or wrong depending on whether his response is correct or incorrect. In the following, we will use the word "reinforcement" in the general sense intended by Estes (1960), viz., as operations exerting certain general quantitative effects upon response probabilities. In paired-associate experiments, the reinforcing operation is that of informing subjects of the correct response or in some way ensuring that this response is the last one to occur in the presence of a stimulus item. With this interpretation, then, on every trial under a correction procedure the correct response is reinforced. The noncorrection training procedure differs in that the correct response is reinforced only on those trials when it occurs; since all incorrect responses are followed by the experimenter simply saying "Wrong," no explicit reinforcement operation is involved. We will refer to such trials as nonreinforced trials.

Within the model, an account of the noncorrection procedure takes the form of specific assumptions about the formation of associations following reinforced and nonreinforced trials. Considerations of parsimony lead us to assume that the effect on conditioning of a reinforced trial is the same whether the reinforcement occurs in the context of a correction or noncorrection procedure, viz., the probability that an unlearned item becomes conditioned to its reinforced response is c, the same constant as in the case of the correction procedure. A new assumption that is required concerns the effect of a nonreinforced trial, when the experimenter says "Wrong" following an error. The assumption that has been made elsewhere (Thorndike, 1932) and that is made here also, is that a nonreinforced trial in this situation results in no net change in the probability of a correct response. This assumption is not offered as a universal interpretation of nonreinforcement; however, within the specific situation to which the model will be applied (i.e., the alternative responses are equiprobable initially and a number of different S-R events occur between successive presentations of any given stimulus) there is some basis for thinking that the no change assumption may be approximately correct.

From these assumptions about the noncorrection procedure, a difference equation can be derived expressing the change in the average probability of an error from Trial n to Trial $n+1$:

$$q_{n+1} = q_n[q_n] + (1 - q_n)[(1 - c)q_n] \quad [2]$$

The first term on the right-hand side of Equation 2 exhibits the assumption of no change (i.e., $q_{n+1} = q_n$) when the subject makes an error on Trial n, with probability q_n; the second term shows that if a success occurs (with probability $1 - q_n$), then with probability c the item becomes conditioned (if it is not already) and with probability $1 - c$ the reinforcement was ineffective and no change occurs on that particular trial. An explicit solution of Equation 2 is not available. However, Miller and McGill (1952; see also Bush & Mosteller, 1955, p. 181), have derived the following recurrence relation for Equation 2 (the constants have been changed appropriately):

$$q_n = (1 - 1/N)\,[1 - (1 - c)^{n-1}]q_{n-1} + (1 - 1/N)(1 - c)^{n-1} \quad [3]$$

Equation 3 can be used to compute successive values of the average error probability once N and c are known. In general, the predicted learning curve is **S** shaped, a feature which is consistent with the expectation of a low rate of improvement early in training when the subject is trying to discover the correct response. The point of inflection of the **S** curve will be positively related to N. These features seem in qualitative agreement with noncorrection results reported by Noble (1955), although his experimental procedure was more complex than the one under present consideration.

At this point in the analysis, Experiment I was carried out with four groups of subjects learning the same list of 10 paired associates. The variables were the number of response alternatives (the first three or first eight integers) and correction vs. noncorrection training procedure. For brevity, the four groups will be referred to by the symbols 3-C, 8-C, 3-NC and 8-NC, where 3 or 8 represent the number of responses, and C and NC designate correction and noncorrection training, respectively. A more extensive discussion of the procedure and results of Experiment I will be deferred till later; it will be sufficient here to note one critical result of Experiment I since it was this fact that led to Experiment II. The fact was that strong evidence was obtained to support the assumption of no change in success probability following a nonreinforced trial in the noncorrection training groups; that is, with either three or eight response alternatives, saying "Wrong" followed a subject's incorrect response had no effect on his probability of success on the next trial.

This no change result supports our earlier assumption about nonreinforced trials, but how are we to understand it? On rational grounds, one might expect that subjects would tend to eliminate a response which was followed by "Wrong." For example, if subjects in the 3-NC group tended to eliminate their first erroneous response, then the probability of the correct response on the next trial should be around one-half or at least greater than one-third. However, the results of Experiment I clearly showed that this increase did not occur, and we seek some explanation for why it did not occur.

An explanation may be found perhaps by attending to the responses evoked in the subject by saying to him "Wrong" after he has responded, e.g., "three." The reason for this inquiry is that, holding to a strict contiguity interpretation for the formation of association, it is these terminal reactions to the stimulus which have an opportunity to become conditioned. Often these implicit reactions to "Wrong" are primarily emotional; in this regard, it may be reported that subjects frequently volunteered the information, "I know I've been getting that item wrong, but I can't remember what number I said last time." Clearly, if the subjects' implicit reactions are primarily emotional, then we may expect no change in the recorded response probabilities on the next trial. However, if subjects implicitly react to "three is wrong" with the response "one or two is correct," when the contiguity interpretation would imply that the probability of response "three" would decrease and the probability of responses "one" and/or "two" would correspondingly increase.

According to this interpretation, the noncorrection procedures in Experiment I did not insure responses of the form, e.g., "one or two is correct" after the experimenter said "Wrong" to "three." For one of the groups in Experiment II, conditions were arranged to insure the occurrence of responses in this form following errors. There were three response alternatives and subjects were instructed that one and only one number was correct for each nonsense syllable. If, for example, the subject responded with "three," instead of saying "Wrong," the experimenter said "one or two is correct." It should be noted that, in a formal sense, the subject gains no more information about the correct response with this procedure than when the experimenter says "Wrong," as in the noncorrection procedure. Comparisons of the results of this procedure with the standard noncorrection procedure are given below.

The second condition run in Experiment II was aimed at a slightly different question. The question was whether there is some more basic way of specifying reinforcement contingencies in this situation rather than merely listing them, correction and noncorrection. The formulation proposed here is that this basic variable is the degree to which the experimenter specifies the correct response following an error by the subject; to describe it in another way, the variable is the size of the subset of alternatives within which the one correct alternative is said to lie. The correction and noncorrection procedures occupy the extreme poles of this dimension; in the correction procedure, the experimenter uniquely specifies the correct response; in the noncorrection procedure, when the experimenter says "Wrong," he implicitly specifies "one of the other $N - 1$ alternatives is correct."[3] To illustrate the construction of intermediate values of this training dimension, suppose there were eight response alternatives and that the subject responds incorrectly with "one" to a given item; then the experimenter might say "three or six is correct," or "six or eight or two or five is correct," and so on. For the second group in Experiment II, with eight responses, the experimenter said two numbers following errors. The subjects were instructed that one of the numbers was correct and that the other number was a distractor, but there were no cues as to which number was the correct one. Of course, if the subject gave the correct number, the experimenter indicated this to him.

---

[3] However, see the related discussion above concerning the subject's reactions to "Wrong."

It is clear that such a continuum of training conditions can be constructed. It is also clear that variations in this training variable should produce graded variations in performance intermediate between the two extremes produced by the correction and noncorrection procedures. To describe some of the factors involved in this prediction, consider the reasons for expecting the partial correction group listed above (call it the 8-P group) to make fewer errors than the 8-NC group from Experiment I.

First, a subject in the 8-P condition is expected to make fewer errors before his first correct response because there is at least some likelihood that the correct association will be formed following an error. Specifically, we assume that before the first correct response the probability that the item becomes conditioned following an error is $c/2$, where c is the same learning constant as before. In general, if the subject is told k alternatives, one of which is correct, then we assume that the probability of the correct association being formed is c/k on each trial before the first success. This formulation is equivalent to assuming that the subject selects at random one of the k possibly correct responses to rehearse on a given trial, with probability $1/k$ he selects the correct response, and with probability c rehearsal of this correct response results in conditioning.

The preceding analysis applies on trials before the first correct response occurs. After the subject has been once informed of the correct response, his probability of recognizing it among the two numbers the experimenter says (following a subsequent error, if any) will be greater than one-half. If the subject recognizes the correct number and then rehearses it, the effect is much the same as a reinforced trial. In general, we may let r represent this recognition probability following the first correct response; because of implicit rehearsal, r is also the probability of a reinforced trial on incorrect trials following the first correct response. Specifically, we assume for the present case that r is unity; that is, following the first correct response, we assume that subject can recognize with Probability 1 the correct number among the two numbers that the experimenter says following an error. If the subject recognizes the correct response and rehearses it, then with probability c the conditioned association is formed.

To summarize the discussion for the partial correction groups, we have assumed that prior to the occurrence of the first correct response the probability of conditioning is $c/2$ on each trial; on the trial of the first correct response and those trials following the probability of conditioning is c. In a later section we suggest a number of ways to test the details of these assumptions.

In the experiments to be described there were six independent groups; if the model and theoretical assumptions are correct, then the data from the six groups can be reproduced after estimating the constant c from one of the groups selected arbitrarily. One can easily recognize the advantages of casting our theoretical assumptions in explicit form within such a model. Not only can we clearly pose the questions of how N and its interactions with training conditions affect error rate, but we have also developed a conceptual framework within which it is possible to get an answer to these questions. The basic learning parameter, *c*, plays a central role in the theory and is not just a curve-fitting constant; for a homogeneous population of subjects with the same learning materials, the underlying theory constrains *c* to be the same for all conditions. If indeed such parameter invariance obtains (i.e., if *c* can be transposed from one condition to the others), then a strong appeal can be made for the simplicity and power of the underlying theory within which *c* derives its meaning.

THE EXPERIMENTS[4]

The subjects, 88 Stanford undergraduates, were required to learn a list of 10 paired associates to a criterion of two successive correct runs through the entire list. The stimulus items were nonsense syllables of 0–5% association value chosen for low intralist similarity. The responses were the first three or first eight integers, paired randomly with the stimuli for each subject. With three response alternatives, two of the three responses were used for three stimuli each and the third response, selected randomly for each subject, was paired with the remaining four stimulus items. With eight response alternatives, all eight numbers were first paired with eight randomly selected syllables for each subject, and then the remaining two syllables were assigned to different but randomly selected numbers for each subject.

The subjects, run individually, were instructed to learn a list of 10 nonsense syllables and their associated number responses, the response alternatives being either one, two, or three (or one, two, . . . , seven, or eight), and that the syllables and numbers had been paired off randomly. They were asked to respond within 2 or 3 seconds after the stimulus card was shown, and to guess a number if they didn't know the answer. The result of this practice is that the exposure time cannot be specified exactly; however, it was noted that the subjects followed instructions and responded within 2 or 3 seconds in the large majority of cases.

For all three training conditions, if the subject gave the correct number to a card, the experimenter repeated it, e.g., "three is correct." The three conditions differed on trials when the subject responded incorrectly: in the correction conditions, the experimenter said the correct number; in the noncorrection conditions, the experimenter said "Wrong" following errors; and in the partial correction condition, the experimenter said two numbers following errors, e.g., "one or three is correct." The subjects in the partial conditions were informed that only one of the numbers the experimenter said would be correct and the other number was a distractor (chosen at random) and that half the time the correct number would be said first and half the time it would be said second. The 10 stimulus cards were shuffled for 15 seconds between runs through the deck. The entire list was repeated until the subject went through two consecutive cycles without any errors.

The numbers of subjects in each group were 14 in the 3-C, 8-C, and 8-NC groups; 15 in the 3-P and 8-P groups; and 16 in the 3-NC group. Assuming that each item and each subject may be characterized by the same value of *c*, the data consist of 140, 150, or 160 sequences of correct and incorrect responses for the various experimental conditions. Before the model could be evaluated, the learning rate constant, *c*, had to be estimated. The data from the 3-C group were selected for this purpose;

---

[4] This experiment was carried out with the assistance of Takao Umemoto.

the least squares estimate of $c = .218$ was obtained by fitting the learning curve (i.e., percentage correct vs. trials) for this group.

The first predictions concern the average total errors per item before achieving the learning criterion. The predictions for the correction and noncorrection groups were obtained by summing Equations 1 and 3, respectively, over trials. The predictions for the partial groups were obtained by calculating the average errors before the first success (using $c/2$ as the rate constant) and the average errors following the first success (using c as the rate constant) and then adding these two numbers. The observed means (M) and standard errors of the means (σ_M) are shown in Table 1 along with the predicted means (P).

It is clear from Table 1 that the theory performs adequately in predicting average total errors using the single estimate of c. In all six cases the predicted value falls within one standard error of the observed mean.[5] The differential effect of N upon error rate is amplified as one proceeds from the correction through the partial to the noncorrection training procedures.

TABLE 1

AVERAGE TOTAL NUMBER OF ERRORS PER ITEM OBSERVED AND PREDICTED

Training condition	Response alternatives	
	3	8
Correction	$M = 2.94$ $P = 3.03$ $\sigma_M = .19$	$M = 3.95$ $P = 4.00$ $\sigma_M = .25$
Partial	$M = 3.74$ $P = 3.88$ $\sigma_M = .21$	$M = 5.95$ $P = 5.80$ $\sigma_M = .35$
Noncorrection	$M = 5.65$ $P = 5.59$ $\sigma_M = .34$	$M = 9.42$ $P = 9.54$ $\sigma_M = .61$

A second indication of the goodness of fit of the theory is shown graphically in Figures 1 and 2 which present observed and predicted values for cumulative errors as a function of trials. The smooth curves connect the predicted values while the unconnected points are the empirical values. The predicted values for the correction and noncorrection conditions were obtained by cumulating successive values of error probabilities calculated from Equations 1 and 3, respectively. The theoretical curve for the partial correction conditions is not expressible in a simple equation but successive values can be obtained. The first several values will be derived here to illustrate the procedure. Define Q_n as the probability that the item is not yet conditioned by the beginning of Trial n. The probability of an error on Trial n would then be $(1 - 1/N)Q_n$. The first several values of Q_n for the partial correction conditions are:

by assumption

$$Q_1 = 1$$

$$Q_2 = \frac{1}{N}(1 - c) + \left(1 - \frac{1}{N}\right) \times \left(1 - \frac{c}{2}\right)$$

$$Q_3 = \frac{1}{N}(1 - c)^2 + \left(1 - \frac{1}{N}\right) \times \left(1 - \frac{c}{2}\right)\frac{1}{N}(1 - c) + \left(1 - \frac{1}{N}\right)^2\left(1 - \frac{c}{2}\right)^2 \qquad [4]$$

---

[5] The standard errors for the 3-C and 8-C groups can be predicted by the theory and turn out to be somewhat larger than the observed values (predicted values were .24 and .30 for 3-C and 8-C groups, respectively). Deriving comparable variance predictions for the partial and noncorrection conditions presents complex mathematical problems which have not been solved to date.

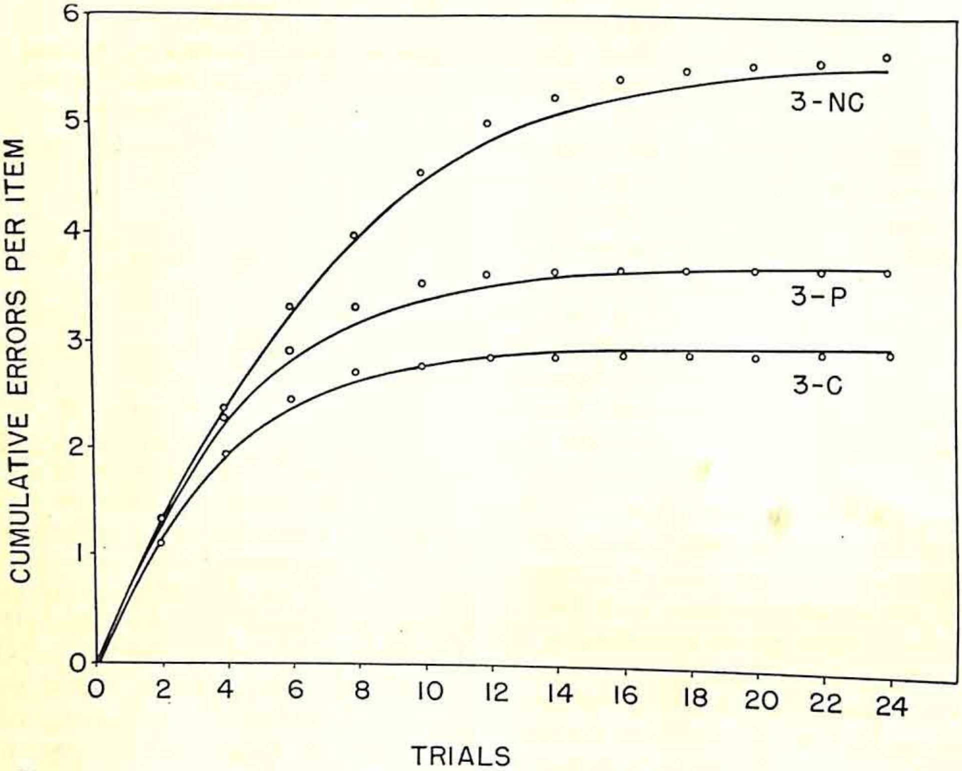


FIG. 1. Average cumulative errors per item plotted as a function of trials for subjects having three response alternatives. (The smooth curves represent predicted values; the dots near a curve represent the corresponding observed values.)

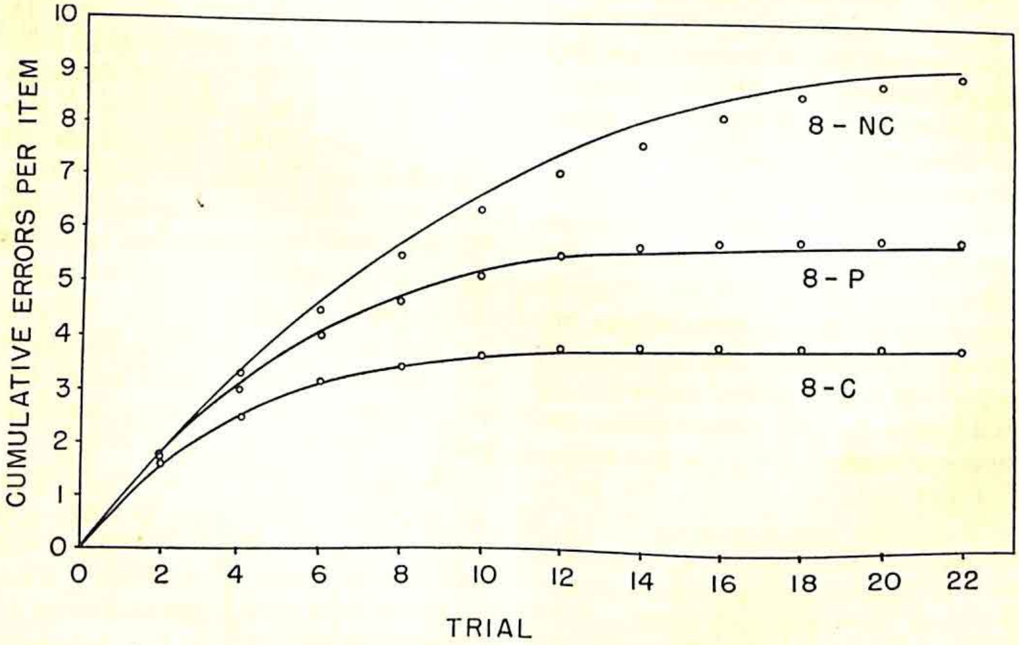


FIG. 2. Average cumulative errors per item plotted against trials for subjects having eight response alternatives. (The smooth curves represent predicted values; the dots near a curve are the corresponding observed values.)

The value of Q_2 is the sum of two probabilities: (*a*) that a correct guess occurs on Trial 1 but conditioning failed to occur with probability $1 - c$, and (*b*) that subject guessed incorrectly on Trial 1 and fails to condition the correct response with probability $1 - c/2$. The value of Q_3 is the sum of the joint probabilities of three events: (*a*) correct on Trial 1, fails to condition on Trials 1 and 2; (*b*) error on Trial 1, success on Trial 2, fails to condition after either trial; (*c*) errors on both trials, not conditioned on either trial. The general rule for obtaining Q_n is that prior to the first correct response the conditioning probability is $c/2$ and that on the trial of the first correct response and thereafter the conditioning probability is c. By multiplying these Q_n values by $1 - 1/N$ and summing them, the predicted curves for the partial correction conditions are obtained. As Figures 1 and 2 show, for five of the six groups the fit of predicted to observed values is satisfactory; the correspondence to the 8-NC points would not be impressive were it not for the fact that the same c constant was used in generating all the curves under appropriate boundary conditions.

Equations 2 and 3 for the noncorrection conditions were derived with the assumption that no change resulted from nonreinforced trials. To the extent that these equations adequately describe performance in the noncorrection conditions, the assumption of no change is supported. However, additional data bearing upon this assumption may be obtained by analysis of the early trials before learning has occurred. Beginning with the first trial, consider the sequence of probabilities of a correct response following 0, 1, 2, . . . consecutive errors. As soon as a subject responds correctly to a given item, that item is dropped from later computations. If the randomization of responses to stimuli has been effective, then the first value of this series should be close to $1/N$, the probability of guessing the correct number by chance. Moreover, if the assumption of no change is correct, then the probabilities of a success following 1, 2, 3, . . . consecutive errors should remain around $1/N$ for the noncorrection groups, deviating from this value only by random sampling fluctuations. The results of this analysis are shown in Table 2 for the noncorrection groups. It should be realized that going down the column these estimates become progressively less reliable because the number of cases involved (beginning with 160 or 140) decreases. These series of estimates for both three and eight response alternatives reveal no significant trends away from the first value. Hence, the results provide additional support for the assumption of no change following nonreinforced trials in the noncorrection conditions.

TABLE 2

PROBABILITY OF A SUCCESS FOLLOWING n CONSECUTIVE ERRORS FROM THE FIRST TRIAL FOR THE NONCORRECTION GROUPS

n	3-NC	8-NC
0	.34	.12
1	.31	.11
2	.28	.14
3	.31	.17
4	.31	.13
5	.37	.16

An analysis related to that above can be given for the correction and partial correction groups. Because the success rate was higher for these groups, there was insufficient data to obtain reliable estimates of single-trial probabilities such as shown in Table 2 for the noncorrection groups. How-

TABLE 3

PREDICTED AND OBSERVED CONDITIONAL PROBABILITIES OF SUCCESS FOLLOWING AN ERROR ON THE PRECEDING TRIAL

Training condition	Response alternatives			
	3		8	
	Observe	Predict	Observe	Predict
Correction	.48	.48	.35	.35
Partial				
Before first success	.35	.35	.21	.22
After first success	.46	.48	.38	.33

ever, this difficulty may be circumvented by considering over all trials the probability of a success given an error on the preceding trial. For the correction group, this probability should be a constant, $c + (1 - c)p_o$ where p_o is the probability of guessing correctly. For the partial correction conditions, the probability of a success following an error should be $c/2 + (1 - c/2)p_o$ before the first success and should be $c + (1 - c)p_o$ after the first success. Such predictions are very sensitive to small changes in p_o. Our efforts to randomly assign responses to stimuli were directed towards making p_o be close to $1/N$ in value. However, with even a moderate-sized sample, small deviations of the actual probability of correct guessing inevitably occur. Hence, a fairer test of the model (rather than a test of the randomization procedure) is provided by predictions based on empirical estimates of p_o. Using the first-trial estimates of p_o (which on the whole were close to $1/N$), the predicted probabilities of success following an error were obtained and are compared with the observed values in Table 3. The correspondence between predicted and observed values is quite satisfactory; we may note that the rather complex assumptions about conditioning in the partial correction conditions receive substantial support from these data.

Another way to test the assumptions of the model involves predicting the number of errors before the first success. Since for the partial and noncorrection groups the stochastic process is not complicated prior to their first success, it is possible to predict both the mean and variance of this statistic for these groups. Define F_1 to be the number of errors before the first success, and p_o to be the probability of a correct guess. Then it can be shown (Bower, 1961) that the mean and variance of F_1 for the noncorrection groups are:

$$M = \frac{1 - p_o}{p_o}, \quad V = \frac{1 - p_o}{p_o^2} \qquad [5]$$

Similarly, for the correction groups, the mean and variance of F_1 are:

$$M = \frac{1 - p_o}{p_o + c(1 - p_o)}, \qquad [6]$$

$$V = M + M^2(1 - 2c)$$

The expressions for the partial correction groups are the same as in Equation 6 except with c replaced by $c/2$.

The predicted values of the mean and standard error of F_1 for the six groups are shown in Table 4 along with the observed means and standard errors. For the noncorrection groups, p_o was estimated as the average of the values in Table 2 since the theory says those values represent random fluctuations around p_o.

Inspection of Table 4 shows that in

the six cases the predicted mean F_1 is within a standard error of the observed mean, and the predictions of the standard errors are also close to the observed standard errors. In general, the average F_1 increases with N and the differential effect of N is amplified under the partial and noncorrection training conditions.

It should be noted that at an empirical level N has its primary influence upon errors before the first success, and has practically no influence upon errors following the first success. The reader may convince himself of this fact by subtracting the mean F_1 values in Table 4 from the mean total error values in Table 1. At first glance, this observation seems to invalidate the model by showing that N only influences the discovery stage but not the fixation stage of learning after the first correct response has occurred. In fact, however, theoretical predictions of average errors following the first success correspond closely to the observed values; of course, this must be so since the theory predicts F_1 and total errors reasonably well. The reason we expect errors following the first success to be relatively constant over N is that, according to the theory, a sampling bias is involved in comparing the three and eight alternative conditions at this point. Errors following the first success will occur only for items whose first correct response came about by guessing, and for these guessed correct items we expect more errors with eight than with three response alternatives. Let g_N represent the probability that with N alternatives the first correct response comes about by guessing and, let e_N represent the average number of subsequent errors for an item whose first correct response occurred by guessing. Then the expression for the average number of errors following the first correct response (call it w_N) is:

$$w_N = g_N e_N + (1 - g_N) \cdot 0$$
$$= g_N e_N \quad [7]$$

That is, a proportion g_N of first correct responses come about by guessing and we expect an average of e_N more errors for these items; a proportion $1 - g_N$ of the first correct responses come about by conditioning on the preceding trial and for these items we expect zero subsequent errors. According to the theory, e_8 will be larger than e_3; however, g_8 will be smaller than g_3; that is, given a first correct response, it is more likely to have occurred by guessing when N is three than when N is eight. Therefore, the fact that the products $g_3 e_3$ and $g_8 e_8$ are approximately equal is neither surprising nor a refutation of the theory.

DISCUSSION

The overall results provide strong support for the validity of the model and the specific hypotheses about ex-

TABLE 4

OBSERVED AND PREDICTED MEAN AND STANDARD ERROR OF THE NUMBER OF ERRORS BEFORE THE FIRST SUCCESS

Training condition	Response alternatives 3	8
Correction	$M = 1.33$ $F = 1.39$ $\sigma_M = .092$ $\sigma_F = .125$	$M = 2.25$ $F = 2.37$ $\sigma_M = .185$ $\sigma_F = .180$
Partial	$M = 2.12$ $F = 2.16$ $\sigma_M = .192$ $\sigma_F = .191$	$M = 4.15$ $F = 4.02$ $\sigma_M = .289$ $\sigma_F = .329$
Noncorrection	$M = 2.17$ $F = 2.15$ $\sigma_M = .201$ $\sigma_F = .206$	$M = 6.20$ $F = 6.38$ $\sigma_M = .511$ $\sigma_F = .558$

perimental variables. The data of six independent groups were adequately reproduced using a single estimate of the learning rate constant obtained from one of these groups. The fact that in 24 independent predictions the model came close to the data is sufficient justification for exploring further consequences of the theory.

The initial problem that led to this investigation was whether the number of response alternatives could be shown to affect learning rate on reinforced trials in addition to contributing to differential error probabilities due to guessing on unlearned items. The data supported the assumption that learning rate on reinforced trials was a constant independent of the number of response alternatives, and that the effect of N upon error rate could be attributed to differential guessing probabilities on unlearned items. It should be noted that these conclusions are valid only for the experimental procedure used here in which the response alternatives are immediately available to the subject. For alternative procedures in which the subject is required to learn the responses (e.g., three or eight nonsense syllables) as well as the S-R associations, a number of complicating factors enter to obscure the picture and prevent resolution of the basic problem.

The second problem of this study was to account for the effect on performance of several training procedures. One may conceive of the correction, partial correction, and noncorrection procedures as varying in the degree to which the correct response is specified following an incorrect response by the subject. This variable in turn determines the probability that the correct response is reinforced following incorrect responses. Thus, in the correction condition, the reinforcement probability is 1 on every trial; under partial correction, the reinforcement probability is $1/k$ before the first correct response and is essentially 1 afterwards due to a high recognition probability; under noncorrection, the probability that the correct response is reinforced is essentially zero following errors. The data supported these hypotheses and gave additional support to the assumption that when the correct response was reinforced, the probability of conditioning was a constant, c, the same for all three conditions.

In addition to confirming these specific assumptions about N and training conditions, the results lend support to the learning model in which these assumptions are embedded. In this one-element model, the probability of the correct response can have only two values, $1/N$ or 1. The stimulus element begins in the unconditioned state and each reinforced trial provides an opportunity for the element to become associated in all-or-none fashion with the correct response. When conditioning finally occurs, response probability jumps from $1/N$ to 1. Discontinuous learning theories have been contrasted frequently with "response strength" theories which assume that an associative factor (habit strength in Hull's theory, proportion of conditioned elements in Estes' linear model) increases in a cumulative manner with successive reinforcements. According to a strength theory, response probability make take on a large (possibly infinite) number of values ranging from $1/N$ up to 1. It is frequently difficult to distinguish this theory from the one-element model since they predict the same average learning curve over a group of subjects and items. Indeed, the present results on average errors could be fit about equally well by the linear model. Since the discussion has been oriented towards the one-element model, perhaps it would be appropriate to record a few extra results which favor this model. For these purposes, let us consider two sequential statistics for which the two models deliver qualitatively different predictions.

The first statistic is the average number of errors (to perfect learning) following an error that occurs on Trial n of the experiment. The one-element model makes the rather counterintuitive prediction that this average number of subsequent errors is a constant independent of the trial number on which the leading error occurs. Thus, if we observe an error on Trial 10 we predict the same number of subsequent errors as if we had observed that error on Trial 1. The point of the matter is that when an error occurs on Trial n we know that the item was not conditioned before Trial n, and we can set the clock back to Trial 1 as far as the model is concerned in predicting future errors on that item. In contrast, the associative strength approach (specifically, the linear model) predicts that the number of errors following an error on Trial n is a decreasing function of n; that is, the greater the number of reinforced trials before a particular error, the higher the associative strength at that time, and hence the fewer the number of subsequent errors before perfect learning.

To obtain a sizable sample on which to test this critical point, data from 48 subjects in another 10-item paired-associate experiment[6] (with $N = 2$) were pooled with the 3-C and 8-C groups of the present experiment. The varying N's will not affect the results on constancy vs. monotone decreasing aspects of errors following an error on Trial n. Using this pool of 760 response sequences (10 items for each of 76 subjects), the distributions of the number of errors following an error occurring on Trial 1, on Trial 2, . . . , on Trial 6 were obtained. The analysis was not carried beyond Trial 6 since the number of cases involved was decreasing so that estimates of means would be more unreliable. The estimates of the average number of errors following an error on Trial n are shown in Figure 3.

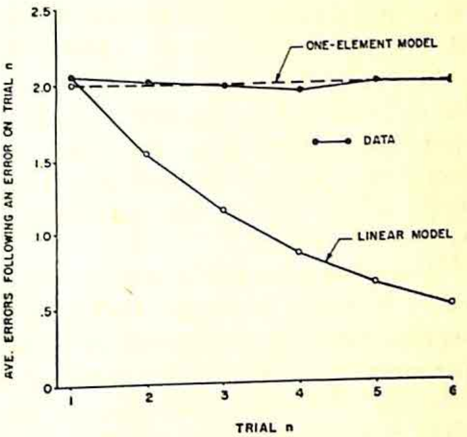


FIG. 3. Paired-associate data: average errors following an error that occurs on Trial n, for $n = 1, 2, \ldots, 6$. (Data points are filled dots; predictions of linear and one-element models are indicated.)

These data clearly favor the constancy prediction of the one-element model. Although the estimates fluctuate somewhat this can be attributed to sampling variability; even the largest difference (2.05 vs. 1.95) does not approach statistical significance ($t = 0.49$). Also in Figure 3 is shown the rough order of magnitude of the numbers to be expected from the linear model. These predicted numbers are not exact because of different learning rates in the two experiments that were pooled. However, the average *c* value was .25 and this value was used to obtain the values on the graph. Clearly the curve predicted by the linear model is quite discrepant from the data; accordingly, the one-element model appears to give a more adequate description of these data.

---

[6] These experiments were carried out with the assistance of Norman Karns and James Colloran.

A second example of this constancy effect is shown in Figure 4 which presents similar data collected in an experiment on verbal discrimination learning. In this experiment 34 subjects learned 20 items, an item consisting of two different nonsense syllables printed on a card. The subject's task was to read off each card that syllable arbitrarily designated as correct for that item; the subject repeatedly went through the items in random order until he achieved two consecutive errorless trials with the entire list. Since response learning per se is not involved in this situation, the one-element model was expected to apply. Since there were two response alternatives on each card, we make the natural assumption that the chance probability that subject selects the correct syllable is one-half. Detailed comparisons of predicted and observed statistics indicated excellent fit of the model; the results in Figure 4 are representative of the overall accuracy of the predictions for this set of data. An additional set of paired-associate learning data providing a comparison of the linear and one-element models has been reported in a previous paper (Bower, 1961); there again the one-element model was clearly superior to the linear model in predicting details of the data.

A second measure suggested by Estes (1960) for differentiating the one-element and linear models con-

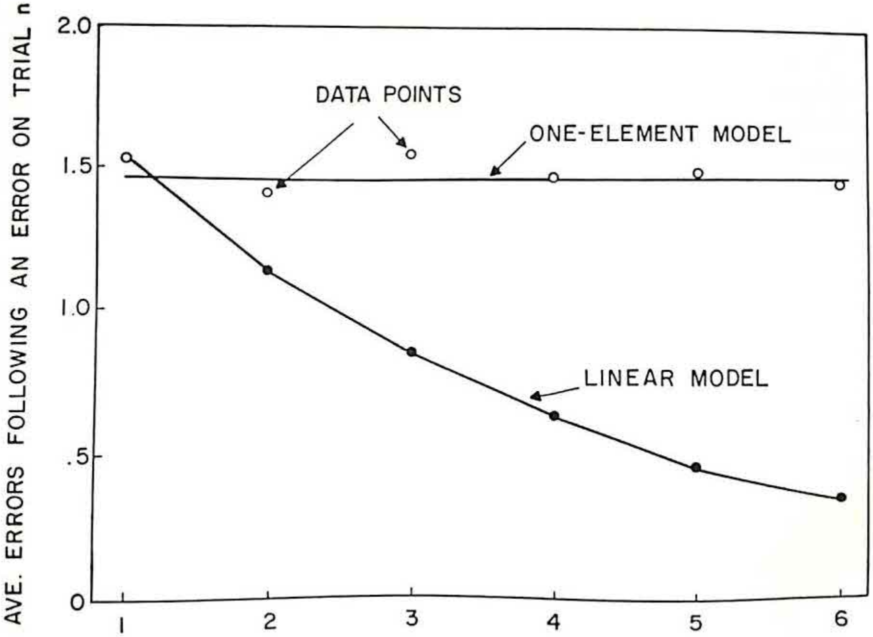


FIG. 4. Verbal discrimination data: average errors following an error that occurs on Trial *n*, for $n = 1, 2, \ldots, 6$. (Data points are the dots: predictions of linear and one-element models are indicated.)

cerns the probabilities of repetition and alternation of responses over a series of test trials following training. Consider a miniature experiment consisting of one or two reinforced trials followed by two test trials (T_1 and T_2) on which the subject is not informed whether his response is right or wrong. The comparison differentiating the two theories is the probability of correct on T_2 given a correct and incorrect response, respectively, on T_1. According to the one-element model, the probability of correct following incorrect on T_1 should be around the chance guessing level ($1/N$); the probability of correct following correct on T_1 will be much higher, being somewhat less than one because we may expect some of the correct responses on T_1 to have occurred by guessing. The predictions of these conditional probabilities by the linear model depends upon whether there are substantial subject and/or item differences in learning rate. If for the moment we neglect such differences, as we have in testing the one-element model, then the linear model implies that the conditional probabilities of success following either a success or failure on T_1 will be equal.

Two experiments will be reported to this point. The first, performed with the assistance of Sharon Gadberry, consisted of 36 subjects learning two 10-item lists (nonsense syllables), the responses being the integers 1 to 10. Following two presentations of each S-R pair, the subjects received two test trials with the stimulus member of each pair. The pooled results of this experiment are shown in the first row of Table 5. Starting with 720 cases (36 subjects on two 10-item lists), the probability of correct on T_1 was .586; of those correct, 89.6% were correct on T_2; of those incorrect on T_1, only 16.3% were correct on T_2. These results support the one-element model. Two comments are required here: First, the value of .896 for correct following correct on T_1 is about what would be expected if 10% of the correct responses on T_1 had occurred by guessing; secondly, the value of .163 for correct following incorrect on T_2 is a little higher than the chance level of .100; however, this is likely an artifact of assigning stimuli and responses in one-to-one correspondence so that when k items are learned the probability of guessing correctly on the unlearned items may be close to one in $N - k$ instead of the a priori value of one in N.

TABLE 5

RESULTS OF TWO EXPERIMENTS ON THE PROBABILITY OF CORRECT ON T_2 FOLLOWING CORRECT AND FAILURE ON T_1

Experiment	Number of cases	Percentage correct on T_1	Percentage correct after correct	Percentage correct after failure
I. Nonsense syllables, 10 number responses	720	.586	.896	.163
II. Noun-noun pair.	320	.900	.995	.030

The difference between probabilities of success on T_2 following a success and failure on T_1 should be increased above those in the previous study if one requires the responses per se to be learned (e.g., as in word-word pairs). Under these circumstances, the chance level of guessing will be practically zero. Thus, if the subject fails to respond correctly on T_1, the probability of being correct on T_2 is essentially zero; by the same reasoning, if a correct response occurs on T_1 it is most likely to have come about by conditioning so that the probability of correct on T_2 for this item should be essentially unity if forgetting is negligible in this situation. Row two of Table 5 reports the results of an experiment performed under these conditions with the assistance of Judith Slagter. Thirty-two subjects received three presentations on each of 10 noun-noun pairs (e.g., moon-pin) followed by two test trials. The percentage correct on T_1 was substantially higher than in the first experiment with nonsense syllable-number pairs. However, the important points of interest are that the probability of correct on T_2 given a correct on T_1 is essentially unity, whereas the probability of correct following failure on T_1 is nearly zero.

One objection frequently raised to these foregoing comparisons is that if there are substantial subject and/or item differences in learning rates, then arguments based upon the pooled aggregate will tend to favor the one-element result. The point of the objection is this: if the linear model holds, and if differences in learning rates lead to a distribution over subjects of response probabilities, then considering the entire aggregate the conditional probability of a success following a success is expected to be higher than the conditional probability of success following a failure. This happens, according to the argument, because when we conditionalize upon the first success we are selecting predominately those protocols from the upper end of the distribution of response probabilities. However, granting that this argument is sound, it nevertheless helps not at all in rescuing the linear model from such data; it simply shifts the focus of the argument from conditional probabilities to joint probabilities of pairs of responses on T_1 and T_2. For convenience in the following, let p_{10} represent the joint probability of a success on T_1 and a failure on T_2, and let p_{01} be the joint probability of failure on T_1 and success on T_2. It is a simple matter to show that if the linear model holds, then p_{10} should equal p_{01} *regardless* of possible differences between subjects and/or items in learning rates or initial probabilities. Data from four studies previously published by Estes (1960, 1961) show that the required identity of p_{10} and p_{01} fails to appear in any of the four studies. For example, in one study (Estes, 1960, Figure 2, p. 215) the value of p_{10} was .132 while p_{01} was .046; in another study (Figure 6, p. 218) p_{10} was .098 while p_{01} was .003. A variety of ad hoc hypotheses about individual differences to supplement the linear model have been considered in a paper by Estes (1961); the outcome of those investigations was that no simple ad hoc supplements could bring the linear model into correspondence with these elementary data which were collected to test it directly.

The question about individual differences raised by the objection above can be answered by comparison with the variability expected from the model. Suppose this question is considered in the context of a more extended experiment such as that described for the 3-C group in the preceding pages. A summary measure of an individual's performance might be the total errors (T) he makes over the 10 items. If we assume within the model that all subjects and items are characterized by the same learning rate constant, c, then a subject's T score represents the sum of 10 values sampled randomly and independently from a distribution which allegedly is the same for all subjects and items. The theoretical variance of these T

scores may be calculated using our single estimate of c; this predicted variance may then be compared with the variance of the observed T scores by taking their ratio which will be distributed approximately as the F statistic. If the observed variance of T scores is much larger than that predicted, then we would tend to attribute the inflated variance to variations in c over subjects. This test has been applied to several sets of paired-associate learning data collected by the writer, including the data from the 3-C and 8-C groups of the present study for which the variance of T could be predicted. With the population of college students who have served as subjects in these various experiments, this test statistic has always yielded nonsignificant F values; that is, the variance of T between subjects does not differ significantly from what would be expected on the basis of random sampling from the stochastic process assumed by the model to be common to all subjects. A similar statement can be made about analogous tests for differences in item difficulty; this outcome was expected, of course, since precautions were taken in selecting stimuli which appeared to be equal in intralist similarity and association value. Alternative procedures for handling this question regarding subject and/or item differences in the RTT experiment were considered in a paper by Estes, Hopkins, and Crothers (1960).

In concluding this report, it might be appropriate to add a few general comments concerning the issues under discussion lest some misunderstandings arise. First, it is misleading to cast the issues under discussion at the level of all-or-none theories versus response-strength theories of learning; rather the point at issue is whether the appropriate model for the present experimental situation is a two-state or a multistate process, where "state" here refers to a particular value of response probability. The continuous linear model is the limiting case of the class of multistate models. The all-or-none assumption about conditioning with respect to the available stimuli is not a differentiating feature of these two classes of models. Since the early writings of Guthrie (1930) it has been clear that the all-or-none conditioning assumption would imply gradual and cumulative changes in response probabilities provided there is sufficient variability in the stimulus samples from trial to trial. Within the framework of statistical learning theory, the feature differentiating the two classes of models is how many independent stimulus components one must assume to represent adequately the course of learning in this or that experimental situation. The two-state model proposed here for elementary association learning assumes that each item may be represented by a single stimulus element within the model and that this element can be in either of two states. That the number of elements is the critical assumption is illustrated by the fact that the statistics in Figures 3 and 4 and in Table 5 would not discriminate qualitatively between the linear model and a *two*-element model (i.e., where each item is represented by two stimulus components, with a random one being sampled on each trial). For more extensive discussions of small element models the reader may refer to a paper by Estes (1959) or a book by Suppes and Atkinson (1960).

Secondly, it should be recognized that in this report we have demonstrated only that the one-element model adequately describes results from elementary paired-associate learning experiments. The job of extending the range of applicability of a theory is an empirical project which must proceed piecemeal. In large part, the success in applying the model to a given learning situation will depend upon the simplicity of the situation and the degree of experimental control over stimulus variables. To cite a pertinent illustration: in training a rat to shuttle in response to a buzzer to avoid shock in a Mowrer-Miller shuttle box, conditioning is typically a gradual, "multi-stage" process (cf. Mowrer, 1960, p. 36); however, if experimental conditions are drastically simplified and precautions are taken to eliminate possible sources of interfering responses, then avoidance conditioning is a one-trial, "two-state" process (Maatsch, 1959). An analogous illustration for the eyelid conditioning situation has been reported by Voeks (1954).

In contrast to those situations in which a two-state behavioral process may possibly be achieved through experimental control, there is a large number of common learning situation for which such precise stimulus control probably can not be achieved, and it is unlikely that a simple two-state model could apply. These situations may be characterized generally as ones in which exposure to discriminative stimuli is subject-controlled; these situations range from those in which the critical cues are proprioceptive stimuli from the subject's current behavior, as in motor skill learning, to situations in which selected components of a complex stimulus array control behavior through the mediation of an overt or implicit observing response, as in complex concept learning. One example of this latter class would be that of a rat learning a black-white discrimination in a T maze; the stimuli effective at the moment of choice are multiple and vary from trial to trial depending upon the subject's vicarious trial and error behavior. For such situations more elaborate conceptual apparatus (cf. Audley, 1960; Bower, 1959; Spence, 1960) is required to analyze the interaction between observing behavior and instrumental, goal-directed behavior.

REFERENCES

AUDLEY, R. J. A stochastic model for individual choice behavior. *Psychol. Rev.,* 1960, **67**, 1–15.

BOWER, G. H. Choice point behavior. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford Univer. Press, 1959.

BOWER, G. H. Application of a model to paired-associate learning. *Psychometrika,* 1961, **26**, 255–280.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford Univer. Press, 1959.

ESTES, W. K. Learning theory and the new mental chemistry. *Psychol. Rev.,* 1960, **67**, 207–223.

ESTES, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika,* 1961, **26**, 73–84.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.,* 1960, **60**, 329–339.

GUTHRIE, E. R. Conditioning as a principle of learning. *Psychol. Rev.,* 1930, **37**, 412–428.

MAATSCH, J. L. Learning and fixation after a single shock trial. *J. comp. physiol. Psychol.,* 1959, **52**, 408–410.

MILLER, G. A., & MCGILL, W. J. A statistical description of verbal learning. *Psychometrika,* 1952, **17**, 369–396.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

NOBLE, C. E. Compound trial-and-error learning as a function of response availability. *J. exp. Psychol.*, 1955, **49**, 93–96.

RILEY, D. A. Rote learning as a function of distribution of practice and the complexity of the situation. *J. exp. Psychol.*, 1952, **43**, 88–95.

SPENCE, K. W. Conceptual models of spatial and non-spatial selective learning. In K. W. Spence (Ed.), *Behavior theory and learning*. Englewood Cliffs, N. J.: Prentice-Hall, 1960.

SUPPES, P., & ATKINSON, R. C. *Markov learning models for multiperson interactions*. Stanford Univer. Press, 1960.

THORNDIKE, E. L. *The fundamentals of learning*. New York: Teacher's College, 1932.

VOEKS, V. W. Acquisition of S-R connections: A test of Hull's and Guthrie's theories. *J. exp. Psychol.*, 1954, **47**, 137–147.

(Received August 23, 1960)

Psychological Review
1962, Vol. 69, No. 1, 54–61

THEORETICAL NOTES

TESTING THE NULL HYPOTHESIS AND THE STRATEGY AND TACTICS OF INVESTIGATING THEORETICAL MODELS

DAVID A. GRANT[1]

University of Wisconsin

Testing the null hypothesis, H_0, against alternatives, H_1, is well established and has a proper place in scientific research. However, this testing procedure, when it is routinely applied to comparing experimental outcomes with outcomes that are quantitatively predicted from a theoretical model, can have unintended results and bizarre implications. This paper first outlines three situations in which testing H_0 has conventionally been done by psychologists. In terms of the probable intentions or strategy of the experimenter testing H_0 turns out to be an appropriate tactic in the first situation, but it is inadequate in the second situation, and it is self-defeating with curious implications in the last situation. Alternatives to this conventional procedure are then presented along with the considerations which make the alternatives preferable to testing the usual H_0.

THREE APPLICATIONS OF H_0 TESTING

Probably the most common application of the tactic of testing H_0 arises when the independent variable has produced a sample difference or set of differences in the magnitude of the dependent variable. Quantitative predictions of the size of the difference or differences are not available. The experimenter wishes to know whether or not differences of the size obtained could have occurred by virtue of the operation of the innumerable nonexperimental factors conventionally designated as random. He sets up H_0 that the differences are zero; chooses a significance level, α; determines the set of hypotheses alternative to H_0 that he is willing to entertain, H_1; selects an appropriate test statistic, t, F, χ^2, U, T, or the like; and proceeds with the test. Rejection of H_0 permits him to assert, with a precisely defined risk of being wrong, that the obtained differences were not the product of chance variation. Failure of the test to permit rejection of H_0, which, unfortunately, is commonly termed "accepting" H_0, means that the obtained differences or greater ones would occur by chance with a probability greater than α. This situation is straightforward. The experimenter has limited aims. He has asked a simple question, and he has received a simple answer, subject only to those ambiguities which attend all experimental and inductive inference. His tactics are admirably suited to his strategic objective.

Another common but less satisfactory instance of testing H_0 arises when the results of pre-experimental matching or pretesting are to be evaluated. Here the experimenter has measured the dependent variable or some related variable before operation of the independent variable, and he devoutly hopes that the experimental and control groups are alike except for random differences. He is now relieved or chagrined, depending upon whether H_0 is "accepted" or "rejected" as a consequence of his test.

---

[1] The author is indebted to Arnold M. Binder of Indiana University whose arguments inspired him to make explicit some of the issues involved in using conventional analysis of variance procedures in testing the adequacy of a theoretical model. As this paper went through various revisions over a period of time the writer is correspondingly indebted to a number of supporting agencies: the Graduate Research Committee and the College of Letters and Science of the University of Wisconsin, the National Science Foundation, and finally to the Department of Psychology of the University of California, Los Angeles, his host during final preparation of the manuscript.

Even if H_0 is accepted his relief is tempered by some uneasiness. He knows that he has not proved, and indeed cannot prove, that H_0 is "true." His tactics in testing H_0 seem to be appropriate to the impossible strategic aim of proving the truth of H_0. Certainly, if he had a more reasonable aim he has adopted inappropriate tactics. Utilizing these tactics, the best he can do is to beat a strategic retreat, and if H_0 is accepted he can perhaps point out that he has used a very powerful test and that if there were real differences they were most likely very small. Although psychologists have never to my knowledge done so, he might be able to go one step further and point out that his testing procedure would reject H_0 a given percentage of the time, say, 90%, if the "true" difference had been as little as, say, one-tenth of an *SD*. This sort of statement of the power of a test is a commonplace in acceptance inspection (Grant, 1952, Ch. 13).

With the advent of more detailed mathematical models in psychology (e.g., Bush, Abelson, & Hyman, 1956; Bush & Estes, 1959; Goldberg, 1958; Kemeny, Snell, & Thompson, 1957) a new statistical testing situation is arising more and more frequently. The specificity of the predictions and perhaps the whole philosophy behind model construction pose a different kind of statistical problem than those faced by most psychological investigators in the past. It seems obvious that as the use of models becomes more widespread a greater number of investigators will face the problem of evaluating the correspondence between empirical data points and precise numerical predictions of these points. Unfortunately most of the procedures used to date in testing the adequacy of such theoretical predictions set rather bad examples. Probably the least adequate of these procedures has been that in which an H_0 of exact correspondence between theoretical and empirical points is tested against H_1 covering any discrepancy between predictions and experimental results.

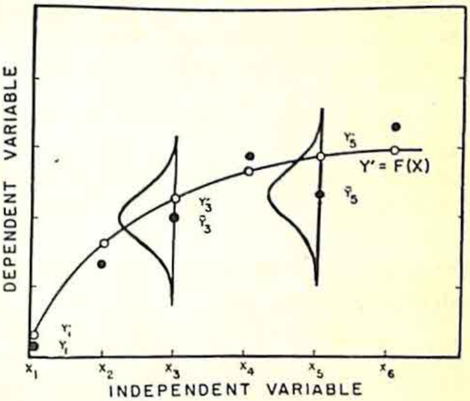


FIG. 1. Idealized situation involving the test of a theoretical function, $Y' = f(X)$. (Theoretical points, Y'_i, are represented by open circles; obtained means, $\bar{Y}_i$, are represented by solid circles.)

Most models predict a considerable number of different aspects of the data, and some of these aspects are predicted with greater success than others (Bush & Estes, 1959, Chs. 14, 15, 17, 18). We shall restrict our discussion to the prediction of values along a curve which might be a learning curve. An idealized version of such a typical situation is presented in Figure 1. Here, the dependent variable, Y, is plotted on the vertical axis against the independent variable, X, on the horizontal axis. The theoretical model has led to an expression, $Y' = f(X)$, giving a set of k theoretical predictions, $Y'_1, Y'_2, \cdots, Y'_k$. The experiment has produced k empirical data points, a set of mean values, $\bar{Y}_1, \bar{Y}_2, \cdots, \bar{Y}_k$, corresponding to the values of the independent variable that were investigated, namely, $X_1, X_2, \cdots, X_k$. Individual observations tend to form normal distributions about each of the $\bar{Y}_i$, and these normal distributions tend to have equal σ's for all data points. In further discussion we shall assume that inaccuracies in the manipulation of the independent variable, X, can be ignored. The problem now is to investigate the goodness of fit of the Y'_i to the $\bar{Y}_i$ or the correspondence between the Y'_i and the $\bar{Y}_i$.

The tactics oriented toward accepting H_0 as corroborating the theory involve breaking down the jth individual observation from the general mean of all of the observations, as follows:

$$Y_{ij} - \bar{Y} = (Y_{ij} - \bar{Y}_i) + (\bar{Y}_i - Y'_i) + (Y'_i - \bar{Y}) \quad [1]$$

where Y_{ij} is the jth observation in the ith normal distribution, and $\bar{Y}$ is the general mean of all observations.

the total sum of squares may then be partitioned as follows:

$$SS_{\text{Tot}} = SS_{\text{Dev Est}} + SS_{\text{Dev Theory}} + SS_{\text{Theory}} \quad [2]$$

where $SS_{\text{Dev Est}}$ is the sum of squares associated with the variation of individual measures from their means, $SS_{\text{Dev Theory}}$ is the sum of squares associated with the systematic departures of empirical data points from the theoretical points, and SS_{Theory} is the sum of squares associated with departures of the theoretical points from the general mean of the whole experiment.

If we suppose that the linear model for the analysis of variance holds, then:

$$Y_{ij} = \mu + T_i + D_i + e_{ij} \quad [3]$$

where μ is the population mean for all Y_{ij} over the specific values of the independent variable, X_i; T_i is the departure of the "true" theoretical value of Y'_i from μ; D_i is the discrepancy of the "true" value of $\bar{Y}_i$ from the true value of Y'_i; and e_{ij} is a random element from a normal distribution with a mean of zero and variance, σ_e^2, for all i.

For a fixed set of X_i the Ts and Ds may be defined so that $\sum T_i = \sum D_i = 0$. Under H$_0$ each $D_i = 0$. Under H$_i$ some $D_i \neq 0$, and the variance of the D_i, $\sigma_D^2 \neq 0$. This last variance may be termed the true variance of the discrepancies from the theory over the particular set of X_i that was investigated.

The foregoing is a conventional analysis of variance model, and the F ratio of the $MS_{\text{Dev Theory}}$ divided by the $MS_{\text{Dev Est}}$ provides an excellent and powerful test of H$_0$ against H$_1$. The number of degrees of freedom for $SS_{\text{Dev Est}}$ will be $k(n - 1)$ where n is the number of observations per data point, and the degrees of freedom for $SS_{\text{Dev Theory}}$ will be $k - n_T$, where n_T is the number of degrees of freedom lost in the process of fitting the model to the data. If this F is significant, we reject H$_0$, concluding that the discrepancies between the $\bar{Y}_i$ and the Y'_i are too great to be accounted for by the observed random variation in the experiment. In this conclusion we accept the 5% or 1% risk implied by our choice of α.

Logical difficulties arise when F fails of significance. H$_0$ remains tenable but is not proved to be correct. A tenable H$_0$ provides some support for the theory but only in the negative sense of failing to provide evidence that the theory is faulty. To assert that accepting the H$_0$ proves that the model provides a satisfactory fit to the data is an inaccurate and misleading statement. We may mean that we are satisfied, but others, especially proponents of other theories, will tend to regard our test as too lenient.

Failure to reject H$_0$, instead of producing closure, leaves certain annoying ambiguities, but the tactics of testing this particular H$_0$ imply a strategy that suffers from more serious defects that are readily apparent when the whole conception of testing a theory is carefully considered. To begin with, in view of our present psychological knowledge and the degree of refinement of available theoretical models it seems certain that even the best and most useful theories are not perfect. This means, in terms of the analysis of variance model, there will be some nonzero D_i's. H$_0$, then, is never really "true." Its "acceptance," rather than "proving" the theory, merely indicates that in this instance the D_i's were too small to be demonstrated by the sensitivity of the experiment in question. The tactics of accepting H$_0$ as proof and rejecting H$_0$ as disproof of a theory lead to the anomalous results that a small-scale, insensitive experiment will most often be interpreted as favoring a theory, whereas a large-scale, sensitive experiment will usually yield results opposed to the theory!

Curiously enough, even rejection of H$_0$ by means of a very stringent experimental test may be quite misleading as far as casting light on the adequacy of the theory is concerned. If the D_i's are very small indeed the theoretical model may be a great improvement over anything else that is available and satisfactory for many purposes even though an extremely sensitive experiment were to reveal the nonzero D_i's. If our task, as scientists, were to test and accept or reject theories as they came off some assembly line the tactics of testing H_0 could be made in a satisfactory manner simply by requiring that the test be "sufficiently" stringent. In fact, our task and our intentions are usually different from testing products; what we are really up to resembles *quality control* rather than *acceptance inspection*, and statistical procedures suitable for the latter are rarely optimal for the former (Grant, 1952, Chs. 1, 13).

HYPOTHESIS TESTING VERSUS STATISTICAL ESTIMATION

An analogy will make clear the relation between testing tactics and the intention of the tester. Suppose that I wish to test a parachute; how should I go about it? How I should test depends upon my general intentions. If I want to sell the parachute and am testing it only to be able to claim that it has been tested, and I do not care what happens to the purchaser, then I should give the parachute a most lenient, nonanalytic test. If, however, I am testing the parachute to be sure of it for my own use, then I should subject it to a very stringent, nonanalytic test. But if I am in the competitive business of manufacturing and selling parachutes, then I should subject it to a searching, analytic test, designed to tell me as much as possible about the locus and cause of any failure in order that I may improve my product and gain a larger share of the parachute market. My contention is that the last situation is the one that is most analogous to that facing the theoretical scientist. He is not accepting or rejecting a finished theory; he is in the long-term business of constructing better versions of the theory. Progress depends upon improvement or providing superior alternatives, and improvement will ordinarily depend upon knowing just how good the model is and exactly where it seems to need alteration. The large D_i's designate the next point of attack in the continuing project of refining the existing model. Therefore attention should be focused upon the various discrepancies between prediction and outcome instead of on the over-all adequacy of the model.

In view of our long-term strategy of improving our theories, our statistical tactics can be greatly improved by shifting emphasis away from over-all hypothesis testing in the direction of statistical estimation. This always holds true when we are concerned with the actual size of one or more differences rather than simply in the existence of differences. For example, in the second instance of hypothesis testing cited at the beginning of this paper, where the investigator tests a pre-experimental difference, he would do better to obtain 95% or 99% confidence interval for the pre-experimental difference. If the interval is small and includes zero, he (and any other moderately sophisticated person) knows immediately that he is on fairly safe ground; but if the interval is large, even though it includes zero, it is immediately apparent that the situation is more uncertain. In both instances H_0 would have been accepted.

TESTING A REVISED H_0

Before turning to estimation procedures that are useful in examining the correspondence between experimental outcomes and predictions from a mathematical model, I shall digress briefly to outline a statistical testing method which can legitimately be used in appraising the fit of a model to data as shown in Figure 1. Basically the statistical argument in the proper test is reoriented so that rejection of H_0 constitutes evidence favoring the theory. The new H_0 is that the correlation between the predicted values, Y'_i, and the obtained values, $\bar{Y}_i$, is zero, after all correlation due to the fitting process has been eliminated. The alternative, H_1, against which H_0 is tested is that there is a correlation greater than zero between theoretical and empirical points. The four simple steps required to obtain the necessary F test are as follows:

1. Calculate $t_i = Y'_i - \bar{Y}$ for all i. $\Sigma t_i = 0$.[2]
2. Calculate $SS_{\text{Correspondence}} = n(\Sigma t_i \bar{Y}_i)^2/\Sigma t_i^2$, where n is the number of observations upon which each $\bar{Y}_i$ is based. Negative values of $(\Sigma t_i \bar{Y}_i)$ are treated as zero.
3. Obtain $MS_{\text{Correspondence}} = SS_{\text{Correspondence}}/n_T$, where n_T, the number of degrees of freedom involved in fitting the theoretical points to the empirical data, will ordinarily be the number of linearly independent fitting constants in the mathematical expression of the model.
4. Divide $MS_{\text{Correspondence}}$ by $MS_{\text{Dev Est}}$ to give $F_{\text{Correspondence}}$ which has n_T degrees of freedom for its numerator and $k(n-1)$ degrees of freedom for its denominator, k being the number of $\bar{Y}_i$. The test is one-tailed in the sense that negative values of $\Sigma t_i \bar{Y}_i$ are treated as zero values, so that the probability values of the F distribution must be halved, an unusual procedure with F tests in analysis of variance.

Following the above procedure, rejection of H_0 now means that there is more than random positive covariation between predicted and obtained values of the dependent variable.

This test is admirable in that it puts the burden of proof on the investigator, because a small-scale, insensitive experiment is unlikely to produce evidence favoring the model. Furthermore, if the model has any merit, the more sensitive the experiment, the more likely it is that a significant F, favoring the theory, will be obtained. Actually, the test is extremely sensitive to virtue in the theory, and therefore in the case of a moderately successful model and a moderately sensitive experiment both this F and the one testing the significance of systematic deviations from the model ($F = MS_{\text{Dev Theory}}/MS_{\text{Dev Est}}$) will tend to be significant. This outcome is no anomaly; it merely indicates that the model predicts some but not all of the systematic variation in the data. In short, progress is being made, but improvement is possible. The fact that simultaneous significance of both Fs, indicating general success and specific failures of a model, should be a commonplace points up the necessity of turning to methods of statistical estimation for a more adequate examination of the workings of a theoretical model.

PRACTICAL ESTIMATION METHODS FOR INVESTIGATION OF MODELS

As is true of statistical tests, each method of statistical estimation has its advantages and limitations. In the investigation of the adequacy of theoretical curves in psychology there are reasons to believe that the simpler estimation methods have practical advantages over some of the more elegant procedures. To give a fairly complete view of the situation, methods of point and interval estimation of σ_D^2 and of the individual D_i will be described, and a brief evaluation of each method will be given.

Estimating σ_D^2. The variance of the discrepancies between the Y'_i and the $\bar{Y}_i$ condenses into a single number the adequacy of fit of the theoretical model. As such it is an excellent index for the evaluation of the model. The smaller the variance, σ_D^2, the better the model, and vice versa. As an estimate of the size of the discrepancies one might expect in future similar applications of the model, σ_D^2 is far more informative than any F test. Furthermore σ_D^2 is readily estimated in the case of homogeneity of the error variance, σ_e^2. The expected values of the relevant mean squares are as follows:

---

[2] In the unusual event where the general mean of the observations, $\bar{Y}$, is not used as a fitting constant for $Y' = f(X)$, the t_i must be computed as deviations from the mean of all the Y'_i, $\bar{Y}'$. The test will then be insensitive to discrepancies between $\bar{Y}'$ and $\bar{Y}$, and the interpretation will be somewhat equivocal. A separate test of H_0 that $\bar{Y}_{\text{population}}$ equals $\bar{Y}'$, is feasible, but here the experimenter is forced into the illicit posture of seeking to embrace H_0.

$$\mathrm{Exp}(MS_{\mathrm{Dev\ Theory}}) = \sigma_e^2 + n\sigma_D^2 \quad [4]$$

$$\mathrm{Exp}(MS_{\mathrm{Dev\ Est}}) = \sigma_e^2 \quad [5]$$

A maximum likelihood estimate of the variance of the discrepancies, $\hat{\sigma}_D^2$ is then:

$$\hat{\sigma}_D^2 = (MS_{\mathrm{Dev\ Theory}} - MS_{\mathrm{Dev\ Est}})/n \quad [6]$$

The accuracy of this estimator depends upon the number of degrees of freedom associated with $SS_{\mathrm{Dev\ Theory}}$ and $SS_{\mathrm{Dev\ Est}}$. The latter rarely poses any practical problem, but the former, in view of the predilection of psychologists for minimizing the number of data points, is quite critical. This is readily apparent when interval estimation of σ_D^2 is attempted.

Bross (1950) gives a convenient method for accurate approximation of the fiducial interval for σ_D^2, and in this case the fiducial and confidence intervals are essentially equal. The method will be outlined below for the 5% interval.

1. Obtain $\hat{\sigma}_D^2$ from Equation 6, above. (If the estimate is negative or zero, meaningful limits cannot be obtained.)

2. Find:

$$L = \frac{\dfrac{F}{F_{025(k-n_T,\ k[n-1])}} - 1}{\dfrac{F \cdot F_{025(k-n_T,\ \infty)}}{F_{025(k-n_T,\ k[n-1])}} - 1}$$

where:

$$F = MS_{\mathrm{Dev\ Theory}}/MS_{\mathrm{Dev\ Est}}$$

$F_{025(k-n_T,\ k[n-1])}$ is the entry in the 2.5% F table (Pearson & Hartley, 1954) for $n_1 = k - n_T$ and $n_2 = k[n-1]$; and $F_{025(k-n_T,\ \infty)}$ is the entry for $n_1 = k - n_T$ and $n_2 = \infty$.

3. Find:

$$\bar{L} = \frac{F \cdot F_{025(k[n-1],\ k-n_T)} - 1}{\dfrac{F \cdot F_{025(k[n-1],\ k-n_T)}}{F_{025(\infty,\ k-n_T)}} - 1}$$

where $F_{025(k[n-1],\ k-n_T)}$ is the entry in the 2.5% F table for $n_1 = k[n-1]$, and $n_2 = k - n_T$; $F_{025(\infty,\ k-n_T)}$ is the entry for $n_1 = \infty$, and $n_2 = k - n_T$.

4. The upper and lower limits are then $\bar{L}\hat{\sigma}_D^2$ and $L\hat{\sigma}_D^2$, respectively. With less than 15–20 data points these limits will be found to be uncomfortably wide, a fact to bear in mind when designing an experimental test of a theoretical model. For example, in Figure 1, with 6 data points and two degrees of freedom for curve fitting, the limits might plausibly be 0–40, whereas with 14 data points the limits might be 0–12.

Aside from the considerable variability in the estimate of σ_D^2 which can be reduced by increasing the number of data points, there are two other important limitations to the use of estimates of the variance of the discrepancies in evaluating a model. First of all, the population value of σ_D^2 is completely dependent upon the particular values of the independent variable, X, which are chosen for the test of the model. Choice of two different sets of Xs could well lead to two entirely different values of σ_D^2, and both of these values could be perfectly accurate. Secondly, although σ_D^2 gives an over-all index of the adequacy of the model being tested, it condenses so much information into one measure that it does not permit pinpointing the especially large D_i's so that they can be given proper attention in considering revision of the model.

Estimating the D_i. The individual D_i may be estimated as points, or intervals may be established for the D_i, collectively or individually. As before, each method has its good points and its limitations.

Point estimation of the individual D_i's consists simply in comparing the individual data points, the $\bar{Y}_i$, with the fitted curve. It is a crude method, but it has served well in the past and represents the beginning of wisdom. For example, in Figure 1, the model builder might well note that the first three data points lie below the curve and ask himself if there is some special reason for this. He would also note that the greatest discrepancy occurs at $\bar{Y}_5$, where the neighboring discrepancies are in the other direction. The weakness of this simple method lies in the absence of a criterion which will assist the investigator in deciding which discrepancies should be singled out for further attention and which may be disregarded because they are within the range of expected random variation. This defect is remedied by the interval estimation techniques.

Probably the ideal method of interval estimation is that in which intervals are established for the whole curve in one operation by finding the 95% confidence band. The method takes the theoretical curve as a point of departure, and the result is a pair of curves above and below the theoretical curve, which will tend in the case of random variation to contain between them 95% of the data points. Points lying outside the band are immediately suspect; they are the most promising candidates for attention in the next version of the model. There are two practical difficulties with this method. First, homogeneity of the error variance, σ_e^2, over all the X_i is required. And secondly, because errors in estimation of each fitting parameter must be taken into account, for all but the simplest curves (Cornell, 1956, pp. 184–186) the bands may be difficult[3] to obtain. Although the method is elegant, in practice it will rarely represent sufficient improvement over the final method, given below, to justify its use.

The last method seems to me to be the most useful and most robust and most flexible method. It can be widely applied, and the relative ease of application, coupled with its ability to discriminate between significant and random discrepancies make it superior to the other estimation methods. It also possesses the homely virtue of being readily understood. In contrast to the preceding method, this one takes as its point of departure the empirical means, and consists, simply, in computing the 95% confidence limits for each of the $\bar{Y}_i$. If there is homogeneity of variance, the error variance of each mean is taken simply as $\hat{\sigma}_e^2/n$; in cases of suspected heterogeneity, each mean must have its own estimate of error variance. This will, of course, be the variance of the distribution of Y_{ij} for each i, divided by n. When these limits have been obtained, attention is directed to instances where the theoretical curve lies outside the limits. In some cases, the investigator might choose to establish the 80% or 90% limits in order to direct his attention to less drastic departures of the experimental results from the model. Choice of an optimum level for the limits is hard to establish on a general a priori basis, but it is likely that limits narrower than the traditional 95% will be found more useful than the broader limits. Simple as this method is, it is hard to improve upon in actual practice. Instead of giving an almost meaningless over-all acceptance or rejection of a model, it directs attention to specific defects, its functioning improves as the precision of the experimental test is improved, and the investigator can set the confidence coefficient so as to increase its sensitivity to defect at a cost of a fairly well-specified percentage of false positives or wild goose chases. A final and often crucial advantage is that the confidence intervals, based as they are upon the experimental means, can be obtained in cases where the form of the theoretical function does not permit satisfactory estimation of its parameters, and the analysis of variance and confidence bands methods cannot properly be applied.

SUMMARY AND CONCLUSIONS

In this paper I have attempted to show that the traditional procedure of testing a null hypothesis (H_0) of a zero difference or set of zero differences is quite appropriate to the experimenter's intentions or scientific strategy when he is unable to predict differences of a specified size. When theory or other circumstances permit the prediction of differences of specified size, using these predictions as the values in H_0 is tactically inappropriate, frustrating and self-defeating. This is particularly true when a theoretical curve has been predicted, and H_0 is framed in terms of zero discrepancies from the curve. If rejection of H_0 is interpreted as evidence against the theory, and "acceptance" of H_0 is interpreted as evidence favoring the theory, we find that the larger and more sensitive the experiment is, the more likely it will lead to results opposed to the theory; whereas the smaller and less sensitive the experiment, the more likely the results will favor the theory. Aside from this anomaly, which can be corrected by recasting H_0 in terms of a zero covariance between theoretical prediction and experimental outcome, hypothesis testing as a statistical tactic in this case implies an acceptance-inspection strategy. Acceptance-inspection properly involves examination of finished products with a view to accepting them if they are good enough and rejecting them if they are shoddy enough. The theoretician is not a purchaser but rather he is a producer of goods in a competitive market so that his examination of his theory should be from the standpoint of quality control. His idealized intentions are to detect and correct defects, if possible, so that he can produce a more adequate, more general theoretical model. Because his ideal strategy is not to prove or disprove a theory but rather to seek a better theory, his appropriate statistical tactics should be those involving estimation rather than hypothesis testing.

---

[3] A sufficient estimate of the error variance of each parameter must be available and independent of the estimates of all other parameters or else the covariances of all parametric estimates must be found and the theoretical function must have continuous first partial derivatives with respect to the parameters in order that the confidence bands may be found in the asymptotic case (Rao, 1952, pp. 207–208). Where an asymptote is involved in the fitting of the theoretical function, satisfactory independent estimators can rarely be obtained.

Examination of alternative techniques available for point or interval estimation of discrepancies between theoretical predictions and experimental outcomes or the over-all variance of these discrepancies suggests strongly that estimation of the confidence intervals for the means found along a theoretical curve is the most practical and most widely applicable general procedure. Other writers have recently emphasized the values of various estimation as opposed to hypothesis testing techniques (e.g., Bolles & Messick, 1958; Gaito, 1958; Savage, 1957) and it is hoped that considerations pointed out by them and points raised in this paper will be helpful to investigators who are in the process of examining theoretical models which lead to specific numerical predictions of experimental outcomes.

REFERENCES

BOLLES, R., & MESSICK, S. Statistical utility in experimental inference. *Psychol. Rep.*, 1958, 4, 223–227.

BROSS, I. Fiducial intervals for variance components. *Biometrics*, 1950, 6, 136–144.

BUSH, R. R., ABELSON, R. P., & HYMAN, R. *Mathematics for psychologists: Examples and problems.* New York: Social Science Research Council, 1956.

BUSH, R. R., & ESTES, W. K. (Eds.) *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959.

CORNELL, F. G. *The essentials of educational statistics.* New York: Wiley, 1956.

GAITO, J. The Bolles-Messick coefficient of utility. *Psychol. Rep.*, 1958, 4, 595–598.

GOLDBERG, S. *Introduction to difference equations.* New York: Wiley, 1958.

GRANT, E. L. *Statistical quality control.* New York: McGraw-Hill, 1952.

KEMENY, J. G., SNELL, J. L., & THOMPSON, G. L. *Introduction to finite mathematics.* Englewood Cliffs, N. J.: Prentice-Hall, 1957.

PEARSON, E. S., & HARTLEY, H. O. (Eds.) *Biometrika tables for statisticians.* Cambridge, England: Cambridge Univer. Press, 1954.

RAO, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

SAVAGE, I. R. Nonparametric statistics. *J. Amer. Statist. Ass.*, 1957, 52, 331–344.

(Received November 10, 1960)

Psychological Review
1962, Vol. 69, No. 1, 62–64

THE CONSTRUCTION OF *Q* SORTS:

A CRITICISM

DONALD M. SUNDLAND[1]

Columbus Psychiatric Clinic

The widespread use of *Q* sorts makes it important that certain difficulties in item selection be emphasized. In this author's opinion, these difficulties are such as to suggest the discontinuance of the use of this technique for research purposes, unless corrective measures are devised.

The *Q* sort is a technique derived from *Q*-methodology practice (Mowrer, 1953; Stephenson, 1955). In using this technique, comparison is made between answers to a large number of "tests" (items) from different persons (usually judges), or from the same person from different points of view: "self, ideal self, self as seen by mother, etc." An implicit, underlying assumption is that the tests are not inherently correlated. Operationally speaking, if two judges agree that the quality "shy" is "very characteristic" of Person X, then their agreement about the qualities "bashful, modest, outgoing, extroverted, etc." is, at least partially, predetermined by the inherent correlations between the items. Now this would be all right if all the items were about shyness or nonshyness, then agreement between judges would mean that they agreed on the shyness of X. Characteristically this is not the case. Items relative to the broad range of personality traits are usually included in a *Q* sort. Under these circumstances the inclusion of items inherently related will lead to spurious correlations between *Q* sorts.

This difficulty has been pointed to before, but seemingly without appreciation of its seriousness. Stephenson (1955) wrote: "no set contained direct antonyms" (p. 257) suggesting that this was something to be avoided. Later Stephenson[2] said:

> you are quite right about . . . the use of direct antonyms. Little is to be gained by using words like introversive and extroversive in the same *Q*-sample, one of them is enough.

Cronbach (1953) said, in discussing criteria for the selection of items for *Q* sorts:

> statements, while logically bearing on the same domain, should represent a large number of continua. Correlating persons seems to have no advantage if we present items which all fall into one scale (p. 380).

Two types of statistical distortion are involved in this use of highly related items. The first is an inflation of the correlation coefficient, and the second is an inflation of the number of degrees of freedom. By using the total number of presumably unrelated items as the degrees of freedom, one markedly inflates the degrees of freedom in testing the significance of the correlation if many of the items are, in fact, correlated. A rather obvious example would be using only items related to a single dimension. If 30 items concerning a single trait were used to judge a person, the correlation between judges would be accurate and misleading at the same time. It would be accurate insofar as the judges' agreement about the trait, but it would be misleading about the person insofar as this is the only trait being considered. This is pretty obvious; however, now let us intermingle these 30 monotrait items among 70 other items, these 100 items supposedly representing the broad spectrum of personality traits. Further, let

---

[1] The author expresses his thanks to Robert J. Wherry, Edwin N. Barker, and Shephard Liverant for their valuable comments on this paper.

[2] W. Stephenson, personal communication, 1957.

us suppose that two judges have a zero correlation on these 70 other items, but they have a perfect, 1.00, correlation on the 30 monotrait items. The result of this intermingling would depend upon the placement of the 30, agreed upon items. The author devised a set of data to examine this problem. Seventy items were randomly distributed into the 11-category, approximately normal distribution shown in Table 1 and then minor adjustments were made to obtain an exactly zero correlation. The 30 agreed upon items were then intermingled in three ways. The first way was to place them in the extreme categories of the combined 100-item distribution shown in Table 1; the second was to place them in the middle categories; and the third was to distribute them proportionately. This procedure yields the correlations: 0.76, 0.11, and 0.30. A correlation of 0.26 for 100 items is significant at the 0.01 level. Considering the monotrait items as really one item, this example shows two types of overestimation: the correlations found range between .11 and .76 instead of from .00 to .06, and the degrees of freedom used are 98 (100–2) instead of 69 (71–2).

Let us look at another frequent use of the *Q* sort: self vs. ideal-self comparison. Where the monotrait items are rated in the same direction on both self and ideal-self *Q* sorts, the correlation will be overestimated. Where the monotrait items are rated in the opposite directions, the correlation will be underestimated. This distortion would be augmented if the experimenter were comparing *Q* sorts before and after therapy and if the monotrait items shifted their directions. For example, let us suppose that before therapy a patient sees himself as shy but as ideally not shy, and that after therapy he sees himself as not shy both actually and ideally. If the *Q* sort had many "shyness" items in it, the difference between the pretherapy and posttherapy correlation would be doubly overestimated since the pretherapy correlation would have been underestimated and the posttherapy correlation would have been overestimated.

TABLE 1

ITEM DISTRIBUTIONS

Number of Items	Categories										
	1	2	3	4	5	6	7	8	9	10	11
70 items	2	5	5	7	7	18	7	7	5	5	2
100 items	3	7	7	10	10	26	10	10	7	7	3

It would seem that a possible hope for the comparison of persons, or of different concepts of a person, would be the use of items whose interrelatedness is known. Levy and Dugan (1960) have proposed a method which offers a possible system for determining the interrelatedness of items. An investigator adapting their design would ask the subjects, from the same population as his experimental subjects, to describe a photograph with his *Q*-sort items. Each subject would be given a different photograph. The items could then be intercorrelated and factor analysed into orthogonal components. In using the *Q* sort with his experimental subjects the investigator could then obtain factor scores and use these for his comparisons. Since this is a somewhat laborious procedure it would seem to be desirable for investigators to come to some consensus about the items for two or three widely applicable *Q* sorts. Short of the availability of independent items or items with known dependencies, any kind of *Q*-sort study is subject to this distortion.

SUMMARY

It has been pointed out that due to the interrelatedness between items which constitute a *Q* sort, investigators obtain spurious results due to an overestimation of the correlation coefficient and of the degrees of freedom in their *Q* sort. The difficulty of determining the correct magnitude of the correlation coefficient and the correct number of degrees of freedom to be used indicates the desirability for the development of alternate methods for Q-type studies.

REFERENCES

CRONBACH, L. J. Correlations between persons as a research goal. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research.* New York: Ronald, 1953.

LEVY, L. H., & DUGAN, R. D. A constant error approach to the study of dimensions of social perception. *J. abnorm. soc. psychol.,* 1960, **61**, 21–24.

MOWRER, O. H. *Q* technique-description, history and critique. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research.* New York: Ronald, 1953.

STEPHENSON, W. *The study of behavior.* Chicago: Univer. Chicago Press, 1955.

(Received December 1, 1960)

VOL. 69, NO. 2 MARCH 1962

PSYCHOLOGICAL REVIEW

TWO COMPONENTS IN BIPOLAR SCALES: DIRECTION AND EXTREMENESS[1]

DEAN PEABODY[2]

Columbia University

Various measures in psychology use a set of response categories extending in two directions about some neutral point. Scores for several responses are summed or averaged to give a composite score. Examples would include some of the methods of absolute judgment in psychophysics, the Likert-type of attitude scale, Osgood's semantic differential, rating scales used in personality assessment. The object of measurement for the composite scores may thus be some attribute of physical stimuli, of verbal concepts, of other persons, or of oneself.

Two components can be distinguished in the responses: (*a*) the direction from the neutral point, representing the basic dichotomy—heavy versus light, agreement versus disagreement with the statement, etc.; (*b*) the degree or extremeness of the response from the neutral point. For example, in a six-point Likert scale the subject writes + 1, + 2, + 3; or − 1, − 2, − 3 to indicate slight, moderate, and strong agreement or slight, moderate, and strong disagreement. The basic dichotomy of direction is determined by whether he agrees or disagrees (writes + or −); extremeness by whether he indicates slight, moderate, or strong feelings (writes 1, 2, or 3).

A neutral response category may or may not be provided—e.g., Likert scales traditionally permitted a neutral response (writing 0), but many recent Likert scales do not.

When scores for several responses are summed or averaged to give a composite score, the two components become: the number (more generally, the proportion) of responses scored in each direction; the mean extremeness of the responses in each direction.

The present paper derives a method for estimating the relative contribution of these two components to composite scores, and applies the method to data from several Likert attitude scales.

A possible reaction to the preceding might be to grant that the two components described can logically be distinguished, but question whether this is necessary or desirable. The purpose of using an extended scale is to increase the refinement of measurement over that given by dichotomous responses. What then is the purpose of separating direction and extremeness components? The answer is that there are both theoretical and practical reasons for considering the components separately:

---

[1] This report represents a revision of a doctoral dissertation at Harvard University. The collection of data in England was aided by a grant from the Social Science Research Council; the replication in the United States by a grant from the Laboratory of Social Relations at Harvard. Support for rewriting was given by the Council for Social Sciences at Columbia. I am particularly grateful to G. W. Allport and F. Mosteller for general and statistical advice.

[2] Now at the University of Oregon.

1. The theoretical reason is that the distinction may be a psychological as well as a logical one. Ideally, the direction component would represent a qualitative aspect of the object of measurement, while the extremeness component would represent quantitative, intensity, aspects. In reality, both components are likely to represent, at least in part, response tendencies of the subjects rather than aspects of the object of measurement. Cronbach (1946; 1950, p. 21) distinguishes the response sets possibly involved in the two components as "bias" or "acquiescence" for the direction component, and "tendency to go to extremes" for the extremeness component. It seems desirable to examine these different factors separately. For example, Cronbach (1950) argues strongly that the extremeness component may largely represent a response set:

> The usual argument for the more finely divided scale of judgment on each attitude item is that it is more reliable and that subjects prefer it. If the latter advantage is significant, the finer scale may be retained and scored dichotomously. The argument that the finer scale gives more reliability is not a sound one, since this is precisely what we would expect if all of the added reliable variance were response-set variance and had no relation to beliefs about the attitude-object in question (p. 22).

2. The practical reason for examining the two components separately concerns the justification for dichotomous scoring. Regardless of the theoretical status of the two components, if the extremeness component is little represented in whatever the composite score measures, use of the simpler scoring method based on the direction dichotomy would seem justified.

DEVELOPMENT OF AN APPROXIMATION FOR THE VARIANCE OF COMPOSITE SCORES IN TERMS OF DIRECTION AND EXTREMENESS SCORES

The development given below assumes that a neutral response category is not provided for. Provision for a neutral category would give a closely analogous development.

The composite score used here represents the average of scores for the separate responses. An individual's composite score can be expressed mathematically in terms of direction and extremeness scores:

$$C = PX - (1 - P)\ Y \qquad [1]$$

where:

C is the composite score.
P is the proportion of responses in the direction scored positively.[3]
X is the mean extremeness of responses in the direction scored positively.
Y is the mean extremeness of responses in the direction scored negatively.

For example, a subject makes 10 responses as follows: $+2$, $+2$, $+2$, $+2$, $+3$, $+1$, -1, -1, -2, -3. If the responses are scored as they stand, then $P = .60$, $X = 2.00$, $Y = 1.75$, and $C = +.50$.

Frequently, to avoid negative numbers some constant is added to all scores. This simply raises the composite score and does not affect the development given below.

The present problem now becomes: to what extent does variation in the composite scores (C) depend on direction (P), and to what extent on extremeness (X and Y)?

---

[3] If a neutral response category were allowed, then Equation 1 should be modified to replace $1 - P$ by $Q =$ the proportion of responses in the direction scored negatively.

The variance of a difference in this case can be written as:

$$\sigma_C^2 = \sigma_{PX}^2 + \sigma^2_{(1-P)Y} - 2\ cov\ [PX,\ (1-P)Y] \qquad [2]$$

The terms on the right involve the variances and covariance of products. There seems to be no simple method of handling this situation. The method used here is to approximate the terms on the right side of Equation 2 by the delta method.[4] In the present case, the assumption is that the deviations of individual scores around the group means are relatively small compared to the means, so that higher powers of the deviations can be neglected.

Let M_P, M_X, and M_Y stand for the group means of P, X, and Y. These means act as constants as regards variations between individual scores.

Let Δp, Δx, Δy stand for individual deviations around the group means:

$$P = M_P + \Delta p$$
$$X = M_X + \Delta x$$
$$Y = M_Y + \Delta y$$

These deviations are assumed to be small compared to the means.

The method will be illustrated for the first term on the right of Equation 2, σ_{PX}^2.

By the definition of any variance:

$$\sigma_{PX}^2 = E[(PX)^2] - [E(PX)]^2 \qquad [3]$$

Applying the delta method:

$$PX = M_P M_X + \Delta p M_X + \Delta x M_P + \Delta p \Delta x \qquad [4]$$

In taking expected values, the terms involving the first order of Δ vanish, since they are deviations around a mean, with an expected value of zero. The expected value of $\Delta p \Delta x$ is the covariance $(\Delta p,\ \Delta x)$. Hence:

$$E(PX) = M_P M_X + r_{(P,X)}\sigma_P\sigma_X \qquad [5]$$

In squaring Equation 4 the terms are collected according to the order of Δ involved:

$(PX)^2 =$

0 order of Δ: $M_P^2 M_X^2$

1 order of Δ: $+ 2\Delta p M_P M_X^2 + 2\Delta x M_P^2 M_X$

2 order of Δ: $+ \Delta p^2 M_X^2 + \Delta x^2 M_P^2 + 4\Delta p \Delta x M_P M_X$

$$\left.\begin{array}{l}3\text{ order of }\Delta:\\ 4\text{ order of }\Delta:\end{array}\right\}\begin{array}{l}\text{These terms will}\\ \text{be neglected.}\end{array}$$

In taking expected values, the first order terms again vanish. The third and fourth order terms are assumed to be negligible. The result is:

$$E[(PX)^2] = M_P^2 M_X^2 + \sigma_P^2 M_X^2 + \sigma_X^2 M_P^2 + 4r_{(P,X)}\sigma_P\sigma_X M_P M_X \qquad [6]$$

Substituting from Equations 5 and 6 in Equation 3 gives:

$$\sigma_{PX}^2 = \sigma_P^2 M_X^2 + \sigma_X^2 M_P^2 + 2r_{(P,X)}\sigma_P\sigma_X M_P M_X - r^2_{(P,X)}\sigma_P^2\sigma_X^2 \qquad [7]$$

Equation 7 gives an approximation for the first term on the right of Equation 2 in terms of the means, variances, and covariance of P and X.

For the second term on the right of Equation 2, an analogous application of the delta method gives the approximation:

$$\sigma^2_{(1-P)Y} = \sigma_P^2 M_Y^2 + \sigma_Y^2(1 - M_P)^2 - 2r_{(P,Y)}\sigma_P\sigma_Y(1 - M_P)M_Y - r^2_{(P,Y)}\sigma_P^2\sigma_Y^2 \qquad [8]$$

For the third term on the right of Equation 2, the approximation is:

$$cov\ [PX,\ (1-P)Y] = -\sigma_P^2 M_X M_Y + r_{(X,Y)}\sigma_X\sigma_Y M_P(1 - M_P) - r_{(P,X)}\sigma_P\sigma_X M_P M_Y + r_{(P,Y)}\sigma_P\sigma_Y M_X(1 - M_P) + r_{(P,X)}r_{(P,Y)}\sigma_P^2\sigma_X\sigma_Y \qquad [9]$$

---

[4] I am indebted to F. Mosteller for first working out the derivation leading from Equation 2 to Equation 10.

Substituting in Equation 2 from the approximations 7, 8, and 9 and collecting terms gives as the overall approximation:

$$\begin{aligned}
\sigma_C^2 &= \overset{\text{I}}{\sigma_P^2(M_X + M_Y)^2} + \overset{\text{II}_\text{a}}{\sigma_X^2 M_P^2} \\
&\quad \overset{\text{II}_\text{b}}{+ \sigma_Y^2(1 - M_P)^2} \\
&\quad \overset{\text{II}_\text{c}}{- 2r_{(X,Y)}\sigma_X\sigma_Y M_P(1 - M_P)} \\
&\quad \overset{\text{III}_\text{a}}{+ 2r_{(P,X)}\sigma_P\sigma_X M_P(M_X + M_Y)} \\
&\quad \overset{\text{III}_\text{b}}{- 2r_{(P,Y)}\sigma_P\sigma_Y(1 - M_P)(M_X + M_Y)} \\
&\quad \overset{\text{III}_\text{c}}{- [r_{(P,X)}\sigma_P\sigma_X + r_{(P,Y)}\sigma_P\sigma_Y]^2} \qquad [10]
\end{aligned}$$

Roman numerals are written above the terms to indicate the grouping that is relevant for present purposes: Term I involves variation of P; Terms II_a, II_b, and II_c involve variation of X and Y; Terms III_a, III_b, and III_c involve the covariation of P with X and Y.

(The final Term III_c involves a higher order of the variables than the other terms, and will often be negligible. The correlations, and the standard deviation of proportion scores are necessarily less than 1. The standard deviations of X and Y also tend to be less than 1 if the scale has no more than three steps in each direction—i.e., a scale of seven steps or less. In such cases the final Term III_c tends to be negligible. For scales that extend further, the term might not be negligible.)

The approximation given by Equation 10 can be used to estimate the relative contribution of direction and extremeness scores to the composite scores on different types of bipolar scales. Comparison can be made of the three groups of terms, and in particular, the relative size of Term I to the combined Term II.

APPLICATION TO DATA FROM SIX-STEP LIKERT ATTITUDE SCALES

This section applies the approximation just developed to data from six-step Likert attitude scales. For each item, the subject chooses between three degrees of agreement by marking $+1$, $+2$, $+3$; or three degrees of disagreement by marking -1, -2, -3. Hence X and Y must lie between 1 and 3, while the proportion scores P lie between 0 and 1.

Before taking up actual data, it is possible to get a general idea of the likely results. If one assumes for example that the means of P, X, and Y are near the middle of their possible values, then M_P will be around .5, and M_X and M_Y around 2. In this case:

Term I is about $16\,\sigma_P^2$.
Term II_a is about $.25\,\sigma_X^2$.
Term II_b is about $.25\,\sigma_Y^2$.
Term II_c is about $-.5\,r_{(X,Y)}\sigma_X\sigma_Y$.

Then with the standard deviations still to be evaluated, Term I has an advantage of 64 to 1 over Term II_a or Term II_b. If one assumes further that the σ's are likely to be roughly in the ratio of their possible ranges, then σ_X and σ_Y would be about twice as large as σ_P. This cuts down the advantage of Term I over Term II_a or II_b to 16 to 1, or an advantage of 8 to 1 over both combined. Insofar as the correlation $r_{(X,Y)}$ is positive, the combined Term II would be made even smaller by Term II_c.

While in principle, data from any set of items could be used, two points deserve consideration.

1. If the items were such that subjects tended to respond in the same direction on nearly all items, then

TABLE 1

Basic Data for Direction and Extremeness Scores

Score	Scale							
	F		Dogmatism		(balanced) Conservatism		General Agreement	
	US	UK	US	UK	US	UK	US	UK
Direction: P								
M_P	.45	.50	.45	.50	.59	.45	.58	.62
σ_P	.12	.18	.10	.13	.11	.16	.05	.06
Rel. of P	.56	.78	.51	.67	.50	.84	.75	.89
Extremeness: X & Y								
M_X	1.79	1.81	1.74	1.75	2.15	1.97	2.01	1.96
σ_X	.34	.31	.29	.28	.35	.36	.25	.23
Rel. of X	.66	.57	.62	.53	.80	.71	.94	.88
M_Y	1.96	1.90	1.94	1.83	1.80	1.95	1.95	1.84
σ_Y	.33	.36	.32	.33	.33	.33	.29	.27
Rel. of Y	.64	.70	.71	.67	.66	.64	.92	.91
Correlations								
$r_{(P,X)}$	.08	.48	.11	.22	.40	.11	−.07	.00
$r_{(P,Y)}$	−.20	−.24	−.08	.01	−.12	−.38	−.14	−.16
$r_{(X,Y)}$	.45	.12	.46	.47	.44	.47	.89	.82

either X or Y would be based on few items and would tend to be an unstable measure.[5]

2. To give a representative test, a variety of types of scales should be used. In particular, the test should include both the recent type of scale deriving from *The Authoritarian Personality* where all items are worded in one direction and agreement is always scored positively, and the more traditional type of Likert scale where items are worded in both directions and agreement is sometimes scored positively and sometimes negatively.

Accordingly results are presented for four sets of items:

1. Responses to 28 items from the F Scale of "authoritarianism" (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, pp. 255–257, 260—omitting Items 22 and 30/28

2. Responses to the 40 items of Rokeach's Dogmatism Scale, Form E (Rokeach, 1956, pp. 7–10)

In both of these scales all items are worded in one direction and agreement is always scored positively. Hence scores represent a mixture of attitudes toward Authoritarianism or Dogmatism with an agreement response tendency ("acquiescent response set").

3. Responses to 32 items referring to conservative versus more leftist positions on economic and political issues. Sixteen of the statements present a conservative position and agreement is scored positively. The remaining 16 statements present the same issue in the reverse direction (the two versions are given at different times) and agreement is scored negatively. Hence the combined score should cancel major effects of acquiescent response set and give a fairly pure measure of attitudes toward Conservatism.

4. Responses to 208 items, scored for "General Agreement." The items include originals and reversals of the F and Dog-

---

[5] In the limiting case, if a subject responds to all items in one direction, either his X or his Y score becomes indeterminate.

TABLE 2

CONTRIBUTION TO VARIANCE OF COMPOSITE SCORES BY DIRECTION AND EXTREMENESS SCORES

Terms from Equation 10	Scale							
	F		Dogmatism		(balanced) Conservatism		General Agreement	
	US	UK	US	UK	US	UK	US	UK
I	.193	.439	.138	.213	.195	.398	.039	.056
II_a	.024	.025	.017	.019	.041	.027	.021	.020
II_b	.032	.032	.031	.026	.018	.033	.015	.011
II_c	−.025	−.007	−.021	−.021	−.024	−.028	−.032	−.024
III_a	.011	.100	.011	.028	.072	.023	−.004	.000
III_b	.031	.057	.010	−.001	.015	.085	.007	.008
II combined	.031	.050	.027	.024	.035	.032	.004	.007
III combined	.042	.157	.021	.027	.086	.108	.003	.008
Total: I + II + III	.265	.646	.186	.265	.316	.538	.046	.071
Actual σc^2	.276	.642	.195	.268	.300	.585	.048	.066
% discrepancy	4%	0.7%	5%	1%	5%	8%	2%	7%
% of Total:								
I: Variance of *P*	73	68	75	81	62	74	85	79
II: Variance of *X* & *Y*	12	8	14	9	11	6	10	10
III: Covariance of *P* with *X* & *Y*	16	24	11	10	27	20	6	11

matism items, as well as for 20 items from the Anti-Semitism scale (Adorno et al., 1950, pp. 68–70); and both versions of the Conservatism items just mentioned. Agreement is scored positively for all items. Hence the combined score should cancel major effects of attitude content and give a measure of generalized "acquiescent response set."

The subjects were 88 American and 75 English engineering college students. Results for the American sample are headed "US"; results for the English sample "UK." Further details of procedure are given elsewhere (Peabody, 1961).

Table 1 gives the means, standard deviations, reliabilities,[6] and correlations for the direction and extremeness scores *P*, *X*, and *Y*.

Table 2 gives the evaluation of the terms of the approximation Equation 10. (The final Term III_c is omitted here since of negligible size.)

The closeness of the approximation is given in Table 2 by the "% discrepancy" between the sum of the terms and the actual variance of the composite scores, and ranges from around 1% to 8%. This result is about what would be expected from this type of approximation.

Although the results in Table 2 differ in detail from the rough a priori estimate made of the relative importance of Terms I and II, the final conclusion is about the same. The average result for the two samples on each set of items gives the following composition of the variance of the composite scores:

---

[6] All reliabilities are for split halves separated by a temporal interval, usually 2 weeks.

> About 70%–80% involves variation of direction scores—Term I
>
> About 10% involves variation of extremeness scores—Term II
>
> About 10%–20% involves covariation of direction scores with extremeness scores—Term III

This is a ratio of 7–8 to 1 for the first two terms.

Two main factors seem responsible for this result:

1. From the general form of Equation 10 variation in one component is weighted by the second power of the mean(s) from the other component (i.e., whether a person agrees or disagrees with an item will usually have a larger effect on his composite score than the degree of his agreement or disagreement). As in the a priori estimate, these weights overwhelmingly favor the direction variance term and more than compensate for the smaller size of this variance itself—so that Term I is considerably larger than any of the Terms II.

2. The correlation between extremeness scores in the positive and negative directions is always positive. Those who use large positive numbers (strongly agree) tend to use large negative numbers (strongly disagree). Hence Term II_c cancels much of the effect of the extremeness component. This factor becomes particularly prominent for the General Agreement scores where the variance of direction scores is relatively small (presumably due to the balancing out of content), but the correlation between X and Y becomes very substantial.

A simpler if less rigorous approach is to consider the more practical question: how good a substitute for whatever is measured by the composite scores can be given by direction or by extremeness scores? Table 3 gives the relevant correlations together with the reliabilities of the composite scores.

The correlations given in Table 3 are representative of the values found for a wider variety of scales. The correlations of composite scores with direction scores tend to be around .90 and above. The correlations with the extremeness scores are more variable and substantially lower.

The results show fairly clearly that composite scores on these Likert attitude items reflect primarily the direction of responses, and only to a minor extent their extremeness. The practical conclusion suggested is that there

TABLE 3

DATA FOR COMPOSITE SCORES

Reliabilities and correlations	Scale							
	F		Dogmatism		(balanced) Conservatism		General Agreement	
	US	UK	US	UK	US	UK	US	UK
Reliability of C scores	.69	.82	.62	.68	.66	.86	.77	.66
Correlations of C scores:								
with Direction scores—								
$r_{(C,P)}$	.93	.97	.93	.96	.87	.97	.96	.95
with Extremeness scores—								
$r_{(C,X)}$	.15	.56	.16	.31	.52	.16	.03	.23
$r_{(C,Y)}$	−.41	−.38	−.38	−.15	−.19	−.47	−.18	−.06

is justification for scoring such items dichotomously according to the direction of response. The resulting score will closely reflect whatever the composite score would.

Further applications are necessary to determine whether these conclusions would extend to scales where a neutral category is permitted, to scales with more than six or seven steps, and to scales that do not concern attitudes. As regards more extended scales (e.g., of 9 or 11 steps) the reasoning used with regard to six-point scales would imply that the extremeness component might play a more significant role. However, in view of the prevalence in psychology of scales with no more than six or seven steps, the present results may be widely applicable.

DISCUSSION

Extremeness scores are quite reliable if based on a large enough number of responses, and are to a considerable extent independent of direction scores. But although extremeness scores can be stable measures of something, they are largely unrepresented in the usual composite scores. Hence the hope that Likert scale scores for example should be "influenced by the *degree* to which subjects favor or oppose attitude statements" (Newcomb, 1950, p. 171) is largely unfulfilled in practice.

Extremeness scores will tend not to have the same correlates as the composite scores, but could have correlates of their own. Only a few studies using separate measures of extremeness give some evidence of what these correlates may be (e.g., Osgood, Suci, & Tannenbaum, 1957, pp. 226–236; Soueif, 1958). In particular, the general conclusion that extremeness is little reflected in whatever the composite scores measure applies to the F Scale of "authoritarianism." This fact is worth noting in view of the recurrent attempts to interpret extremeness as "intolerance of ambiguity," implying that extremeness ought to relate to composite scores on the F Scale—most recently Mogar (1960) uses extremeness on the semantic differential as a criterion for composite F Scale scores.

What is the nature of extremeness scores themselves? Two possibilities have already been suggested: extremeness scores may represent actual differences in intensity (e.g., of attitudes), or they may only represent differences in using the response scale (i.e., response set). Differences within an individual in the extremeness of responses to different items presumably represent primarily differences in intensity, although even here the individual may change his usage of the response scale over time.

On the other hand, as regards differences between individuals in average extremeness, a different interpretation is suggested by the correlations between extremeness scores found in the study reported here. Extremeness seems to be a very general individual characteristic: correlations between extremeness scores in both directions for all possible sets of items (including originals and reversals of the F, Dogmatism, Anti-Semitism, and Conservatism items) tend all to be positive, most of them substantially so. Thus, a person who shows relative extremeness in taking an "authoritarian" position when agreeing with F Scale originals or disagreeing with reversals, tends also to show relative extremeness in taking an "anti-authoritarian" position when disagreeing with F Scale originals or agreeing with reversals, etc. It does not seem very plausible to assume that individual differences in the intensity of attitudes show such systematic generality for responses in opposite directions. But the size of the numbers subjects actually write does show such generality. Hence differences in extremeness between persons seem to represent primarily response sets characteristic of the individual—consistent tendencies to use extreme or moderate response categories. While some evidence suggests a secondary tendency for extremeness scores to represent differences in intensity,[7] the primary factor —the very great generality of extremeness scores—seems better interpreted as response set.

SUMMARY

Scores on the bipolar scales in frequent use in psychology can be analyzed into two components: the proportion of responses in each direction of the basic dichotomy, and the extremeness of responses in each direction. An approximation is developed expressing variation in the usual composite scores in terms of variation in the two components of direction and extremeness.

The approximation is applied to four sets of items from Likert attitude scales. The results indicate that composite scores reflect primarily the direction of responses, and only to a minor extent their extremeness. The practical implication is that there is justification for scoring the items dichotomously according to the direction of response.

Since extremeness scores are reliable and are largely unreflected in the usual composite scores, they may have quite different correlates of their own. Individual differences in average extremeness show wide generality across different extremeness scores. This generality extends to responses in opposing directions—suggesting that the differences primarily represent response sets, and only to a secondary degree actual differences in intensity.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.,* 1946, **6**, 475–494.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.,* 1950, **10**, 3–31.

MOGAR, R. E. Three versions of the F scale and performance on the semantic differential. *J. abnorm. soc. Psychol.,* 1960, **60**, 262–265.

NEWCOMB, T. M. *Social psychology.* New York: Dryden, 1950.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PEABODY, D. Attitude content and agreement set in scales of authoritarianism, dogmatism, anti-semitism, and economic conservatism. *J. abnorm. soc. Psychol.,* 1961, **63**, 1–11.

ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.,* 1956, **70**(18, Whole No. 425).

SOUEIF, M. I. Extreme response sets as a measure of intolerance of ambiguity. *Brit. J. Psychol.,* 1958, **49**, 329–334.

(Received August 23, 1960)

---

[7] For example, correlations tend to be positive between P and X, and negative between P and Y (as in Table 1). This indicates some tendency to use higher or lower positions on the response scale as a whole, as would be expected if intensity were a factor. However, this tendency should lead to negative correlations between X and Y, and is in every case overcome by the opposing tendency for extremeness scores to correlate positively.

Psychological Review
1962, Vol. 69, No. 2, 74–90

THE LATERAL HYPOTHALAMIC SYNDROME:

RECOVERY OF FEEDING AND DRINKING AFTER LATERAL HYPOTHALAMIC LESIONS[1]

PHILIP TEITELBAUM AND ALAN N. EPSTEIN[2]

University of Pennsylvania

The brain controls feeding behavior. If the base of the brain is damaged bilaterally in the ventromedial region of the hypothalamus, rats and other mammals overeat and become tremendously obese (Anand, Dua, & Schoenberg, 1955; Brobeck, Tepperman, & Long, 1943; Wheatley, 1944). Damage in the lateral hypothalamus produces the opposite effect: the animals do not eat and will die in the presence of food and water (Anand & Brobeck, 1951; Anand, Dua, & Schoenberg, 1955). Electrical stimulation through chronically implanted electrodes in the medial hypothalamus suppresses feeding, and stimulation in the lateral hypothalamus elicits it (Brügger, 1943; Larsson, 1954; Wyrwicka & Dobrzecka, 1960).

These dramatic and consistent phenomena suggest that a dual hypothalamic mechanism controls food intake. Cessation of eating (satiety) is said to be controlled by a medial neural system whereas active feeding is maintained by a system in the lateral hypothalamus (Anand & Brobeck, 1951).

There have been two major developments since the original description of the lateral starvation phenomenon by Anand and Brobeck in 1951 which suggest a reinterpretation of the functions of the lateral hypothalamic system. First, not only do the animals refuse to eat, but they also do not drink (Teitelbaum & Stellar, 1954). The adipsia was emphasized by Montemurro and Stevenson (1957). They reported three rats with lateral hypothalamic lesions that did not drink water but did eat appreciable quantities of a high-fat diet, maintaining their weight for as long as 55 days when tube-fed with 10% glucose solution or water. Without hydration, the animals ate less and lost weight. This was the first demonstration that lateral hypothalamic lesions in the rat could produce adipsia without aphagia. The more recent finding that hypertonic saline injected directly into the lateral hypothalamus can elicit drinking as well as eating (Epstein, 1960a) is consistent with this.

Second, the refusal to eat and drink produced by these lesions is not always permanent, and when recovery occurs, it follows a regular and orderly course. Teitelbaum and Stellar (1954), working with large lateral hypothalamic lesions that produced both aphagia and adipsia, kept their rats alive by tube-feeding. Without maintenance these animals would have starved to death, but when kept alive they eventually began to eat and drink again. The animals first accepted wet and palatable foods and began to maintain their weight, then drank water, and finally, regulated their weight on ordinary dry food and water, and seemed normal.

---

[1] This work was supported by NSF Grant G-9792 and by USPHS Grant 2 b-5273. A preliminary report was presented at the 127th meeting of the AAAS in New York on December 29, 1960.

The authors are grateful to Eliot Stellar for his generous advice and material support throughout this work. Both he and R. Duncan Luce made valuable criticisms of the manuscript. Malcolm M. Cohen and Jay J. Singer gave valuable technical assistance.

[2] Formerly at the Institute of Neurological Sciences, Medical School, University of Pennsylvania. Now at the Zoological Laboratories, Division of Biology at the university.

We will show that lateral hypothalamic lesions produce a complex set of symptoms—a syndrome of deficits in drinking as well as in feeding behavior from which the animals recover at different rates. In 35 rats (34 females, 1 male), we have produced bilateral lateral hypothalamic lesions of varying sizes using a standard stereotaxic technique (Stellar & Krause, 1954). The coordinates used were: A 6.0, RL 2.0, and 8.0 down from the dural surface. By offering the animals a variety of wet and palatable foods[3] in addition to the ordinary dry Purina laboratory chow pellets and water, and by keeping them alive by tube-feeding if necessary, we have identified four clear-cut stages in the recovery from lateral hypothalamic lesions. The stages always follow the same sequence and are defined by critical behavioral events that occur in succession during the recovery process. Each of the stages has been analyzed experimentally in an attempt to identify the physiological deficits underlying it.

The animal whose record is graphed in Figure 1 had large symmetrical lesions which resulted in a long, slow recovery, thus making it possible to see the events in each stage quite clearly.

STAGES IN RECOVERY

Stage I: Aphagia and Adipsia

The animals refuse all food, wet or dry, and they do not drink any water. Without tube-feeding, they will die. The animal of Figure 1 neither ate nor drank for 19 days, steadily lost weight, and had to be kept alive from the 7th day on by gastric intubation with a liquid diet. The behavior of such a rat is quite characteristic. It is generally inactive, but it is alert and can be easily roused to movement. It walks with a slightly hunched posture, carrying its head low to the ground. It typically ignores food and will actively push away food that is forced upon it. If such a rat gets wet food on its paws or face, it engages in vigorous bilateral shaking of the front paws which somewhat resembles the paw- and face-grooming movements of the normal rat except that there is no grooming with the mouth. When its paws are wet, it wipes them repeatedly on the floor of the cage, and if the face and mouth have food or water on them, it walks along and rubs its chin and the sides of its face on the cage floor. It does not groom itself effectively, allowing its fur to become completely matted and caked with wet food. It actively resists having milk placed in its mouth by a medicine dropper, and it does not swallow the milk once it is there, but rather allows it to dribble out the side of the mouth. Ordinarily a normal rat does not show such behavior. But it does engage in the same paw-waving and wiping, chin-rubbing, poor grooming, their coats grow sleek as their weight rises, often to preoperative levels. Palatability is no longer essential for food intake. If these rats are trained to press a bar that allows a liquid diet to flow directly into the stomach through a chronically implanted gastric tube (Epstein, 1960b), they ingest normal amounts of food and maintain their weight throughout a 6-day test. They regulate their caloric intake normally even when the diet is diluted with water. When required to press the bar three times for each injection of food, they work sufficiently harder to maintain the intake. This regulation of food intake without stimulation of the oropharyngeal receptors has been demonstrated in three rats during this stage and is similar in all essential respects to that seen in normal animals feeding themselves intragastrically (Epstein & Teitelbaum, in press). The animals press the bar because they are hungry, and the taste of the food is no longer necessary to elicit feeding.

---

[3] A liquid diet was prepared by mixing 75 milliliters of whole egg; 125 milliliters of 50% sucrose solution; 240 milliliters of evaporated milk; 0.3 milliliters Poly-Vi-Sol (multiple vitamin preparation of Mead-Johnson Co., Evansville, Indiana); and 30 milliliters Kaopectate (Upjohn Co., Kalamazoo, Michigan). One part 10% formalin in 100 milliliters of diet was added as a preservative. Milk chocolate refers to finely divided Hershey's milk chocolate. A wet-mash was prepared by mixing Purina Laboratory Chow Powder in equal parts with water.

The animals' feeding behavior seems normal, but they do not drink water. However, they can be weaned from the liquid diet to a sweet nonnutritive solution. By mixing the liquid diet each day with an increasing proportion of the sweet fluid, they can be weaned first to a 10% weight by volume (w/v) sucrose solution, and from this to a 0.5 or 0.2% w/v sodium saccharin solution. Thus, for the first time since the lateral hypothalamus was damaged, they are freely drinking nonnutritive fluid while still refusing dry food and water. As their weight begins to drop, they show the hyperactivity of starvation, then they inevitably begin to eat the dry Purina diet and regain their weight. If the saccharin solution is then removed (as indicated in the hatched area labeled 1 starting on the 49th postoperative day in Figure 1), they still refuse to drink water, their food intake drops to zero in 2 or 3 days, they lose weight, and if prolonged sufficiently, will die. In 28 rats brought to this stage of recovery, all were weaned to sweet fluid and all ate dry food to maintain their weight in the process. For every animal, removal of the sweet fluid led to dehydration because it did not accept water and, consequently, stopped eating. Restore the sweet fluid, however, and it is accepted immediately (as seen in Figure 1). They eat dry food shortly thereafter and then regain their weight. In this stage, therefore, the animals do not eat because they do not drink.

The counter-experiment was performed on four rats. The saccharin solution was again removed, but water, its essential component, was supplied by hydrating the animals intragastrically with 2 to 3 milliliters every hour automatically through a chronically implanted stomach tube. As seen in Figure 1 in the hatched area labeled 2 starting on the 64th postoperative day, the animals continue to eat dry food normally when hydrated, even though they no longer have saccharin to drink. Hydration either by drinking sweet solutions or by intragastric injections allows normal feeding. It is clear, then, that in this stage of the lateral hypothalamic syndrome, the refusal to drink water produces dehydration which in turn prevents feeding.

Does dehydration also account for the failure to eat in Stage I? Will these animals eat immediately after operation if they are hydrated? Chronic gastric tubes were inserted in five animals before they were subjected to bilateral lateral hypothalamic lesions. From the time they recovered from the anesthetic, usually the day after operation, they were hydrated

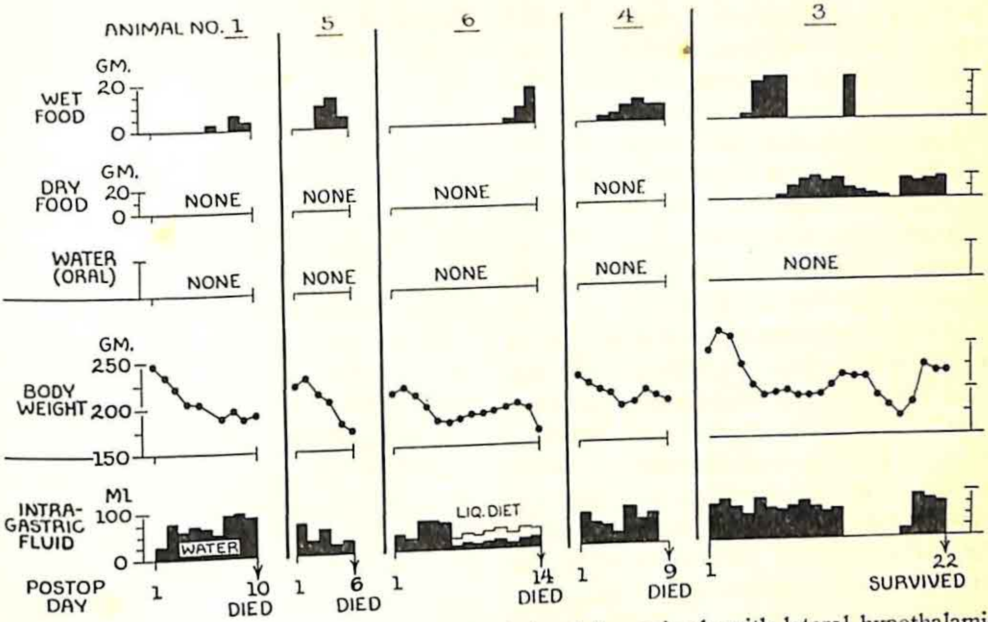


Fig. 2. Food and water intake and body weight of five animals with lateral hypothalamic lesions hydrated automatically from the time of operation. (Four of the animals died after a period of complete aphagia and anorexia despite hydration.)

with 2 to 3 milliliters of water automatically every hour. Figure 2 shows the results. All animals showed complete, immediate postoperative aphagia, varying in duration from 2 to 10 days. Then they all displayed anorexia rather than aphagia, eating some wet food but not enough to maintain their weight. Four of the five animals died. The fifth began to eat dry food on the 7th day and was able to maintain its weight. When hydration was stopped, it stopped eating and lost weight; it ate again immediately after hydration was resumed. It had, therefore, recovered from its initial aphagia and anorexia and passed into Stage III, where it ate if, and only if, it was hydrated. Because all of the animals were initially aphagic and because four of the five animals died despite adequate hydration, we conclude that intragastric hydration initiated in the immediate postoperative period does not assure normal feeding (Morrison & Mayer, 1957). The aphagia and anorexia of Stages I and II are not due just to dehydration.

Why do the rats in Stage III after lateral hypothalamic lesions refuse tap or distilled water and yet readily accept saccharin solutions? Do they drink the sweet saccharin solution because they are thirsty or do they take it in the same way that they take a liquid diet—as a food? A normal animal increases its intake of saccharin solution both when deprived of food (Carper & Polliard, 1953) and when dehydrated. Four normal rats, deprived of food for 3 days, took little or no water but drank increasing amounts of saccharin (from two to four times as much as they did when feeding normally). Three normal rats, dehydrated by an intraperitoneal injection of 2 milliliters of 1 M NaCl per 100 grams body weight (Adolph, Barker, & Hoy, 1954), showed an average intake of 14 milliliters of 0.2% saccharin solution in 2 hours, whereas three other normal rats drank 16 milliliters of water when similarly dehydrated. Thus a normal rat treats a saccharin solution as either a food or a fluid, depending upon its condition of deprivation. When hungry it eats a saccharin solution, and when thirsty it drinks the saccharin.

What does a lateral hypothalamic animal do? When deprived of food for 3 days, four adipsic lateral hypothalamic rats increased their daily intake of saccharin solution (from 20 to 300% more than when feeding freely). However, on repeated occasions, seven adipsic rats drank little or no saccharin solution in response to intraperitoneal hypertonic saline. There were 20 instances of complete lack of response for as long as 5 hours and five instances when they drank only 2 or 3 milliliters. Even a large dose of salt that proved lethal failed to elicit increased fluid intake. Because lateral hypothalamic animals do not increase their saccharin intake when dehydrated but do when deprived of food, they are, in effect, eating the saccharin, not drinking it. They are hungry, not thirsty. Because they do not respond in this stage to levels of dehydration that produce vigorous drinking in normal animals, their refusal to drink water in all three initial stages of recovery may be due to a severe impairment of the central mechanism for water intake regulation.

The refusal to drink water may be permanent, as illustrated in Figure 1. The animal died 200 days after operation, never again having drunk water. It sustained itself for most of that time by "eating" sweet fluid and eating dry food. Three other animals showed permanent adipsia until they died after 267, 230, and 107 days, respectively. One animal, still living, has been adipsic for 485 days.

Stage IV: Recovery

Eventually, after passing through the stages described above, most rats with lateral hypothalamic lesions begin to drink water. Because they continue to eat food and maintain their weight now in the presence of water as the only fluid, they superficially appear normal and in a sense have recovered.

However, more sensitive tests reveal that severe deficits in water regulation still exist. Some animals, for example, still do not respond at all to dehydration by intraperitoneal hypertonic saline. This deficit has been observed for months and may very well be permanent. Also, four of these animals when offered water for only 2 hours each morning did not compensate by taking their daily requirement (approximately 25 milliliters) in that time as normal animals do. They took little or nothing, became progressively dehydrated, stopped eating, and lost weight until they were nearly dead. If they drank at all, they did so shortly after the water was presented. They typically drank 1 or 2 milliliters (never more than 5 milliliters), moved to the back of the cage and slept. When returned to 24-hour access to water, they again managed in an apparently normal manner.

There is a curious paradox here. These animals *do* drink when the water is available overnight but do *not* drink in response to rapid dehydration with hypertonic salt nor do they drink enough when water is available for 2 hours in the morning. The reason for this apparent contradiction was discovered by observing what the animals were doing when they drank. By attaching a drinkometer (Stellar & Hill, 1952) to the water tube and recording the pattern of water ingestion, we found that these animals drank nothing during the day, but drank in short bursts of rapid licking almost continuously throughout the entire night. This is when they normally eat. In addition, if offered both food and water simultaneously at night, they began to drink only after they had begun to eat. Shortly after eating a little dry food, they drank a little water, ate more dry food, drank more water, and so on throughout the night. One of the animals showed this behavior in its most extreme form. It would stand near the drinking tube holding a pellet in its front paws, munch on the pellet for a few minutes, lick actively at the water, and then go back to chewing on the pellet. Its eating was extremely inefficient. It would bite large chunks off each pellet, chew them a little and then drop them while it drank a small quantity of water. In the course of the night it would waste as much if not more food than it actually ate by dropping small chunks of it through the mesh floor of its cage. Its drinking was also abnormal. Only small amounts of water were drunk during bursts of active and precise licking. Although 25 to 30 milliliters of water were consumed, the cumulative record showed continuous licking throughout the night, interrupted by pauses for short bouts of eating. The animal licked all night to consume an amount of water that a normal rat could drink in several sustained draughts. It was as if the animal were drinking not to quench its thirst but simply to wet its mouth.

These observations suggest that such an animal drinks only while it eats. To demonstrate this, it was deprived of water for 24 hours and offered access to it for 2 hours during the night (8 to 10 P.M.) when it normally ate. It drank 22 milliliters of water while eating 12 grams of food. The amount of water deprivation made no difference. It drank only in proportion to the amount it ate. After 48 hours of water deprivation it drank 22 milliliters of water while eating 11 grams of food. When the experiment was repeated in the absence of food so the animal could not eat, it again did not drink any water in 2 hours despite as much as 48 hours of water deprivation. As soon as the food was replaced, the animal started eating and only after eating for a while did it start to drink. The same result was obtained in all four animals that had previously failed to drink adequately in the morning. The animals were then trained to shift their eating into the morning hours (10 A.M. to 12 noon) by offering them food only at that time. Now that they were eating in the morning, they drank enough water in 2 hours to maintain their weight on this regime for several days. Because the animals drink only while they eat, we call this phenomenon "prandial drinking."

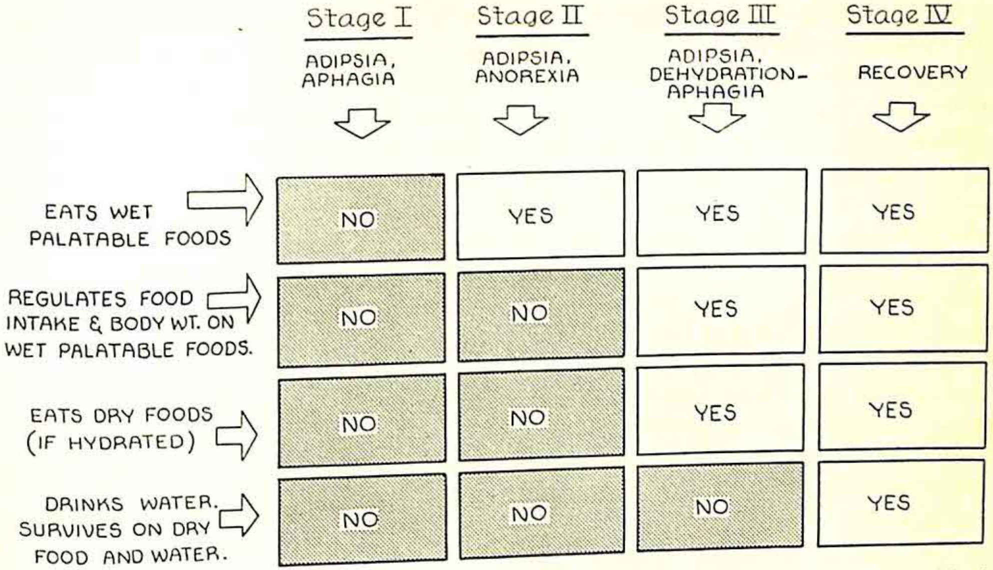
In the same way that the anorexic animals seen in Stage II are eating without caloric regulation, so in Stage IV, when the animals first begin to drink water, they are drinking without response to bodily dehydration. Just as palatability was sufficient to arouse appetite for food, so the mouth factors involved in eating dry food may be sufficient to arouse prandial drinking.

These animals are greatly affected by slight changes in the taste of their water. Normal rats will drink bitter quinine hydrochloride solutions, up to a concentration of 1% w/v, if they are the only fluids available. When ingested at this concentration, 15 to 20 milliliters of quinine hydrochloride is lethal to a rat. A normal rat, therefore, will drink quinine-treated water until it becomes a poison. As this point is approached, they drink less and less, gradually stop eating, and die. Lateral hypothalamic rats that have been drinking water reject quinine hydrochloride solutions at a concentration as low as 0.005%. They refuse it day after day, stop eating, lose weight, and die. But when they are hydrated intragastrically with it, they thrive. Thus, the refusal to eat and drink until death seen in the earlier stages of the lateral hypothalamic syndrome can be reinstated in apparently recovered animals merely by making their water slightly bitter.

A normal rat, as it becomes thirstier, overcomes the barrier to ingestion imposed by the bitter taste of the quinine up to the lethal concentration and drinks enough to stay alive. But a rat with lateral hypothalamic damage does not. Even though it drinks normal amounts and maintains itself when water is constantly available, the lateral hypothalamic animal fails to overcome the quinine taste barrier. In addition, it fails to drink adequate quantities when access to water is limited despite prolonged dehydration, and fails to respond to intraperitoneal saline challenge. These deficits in motivation and in water regulation imply that the central neural thirst mechanism is still severely impaired. The powerful urge to drink is still absent. These animals drink only while they eat, *not* in response to dehydration. There is a striking resemblance here to Adolph's (1957) description of the ontogeny of food and water regulation. The infant rat develops hunger before thirst and, in the neonatal period, does not respond to dehydration. It is adipsic but not aphagic.

Although, according to many measures, food intake has been normal since early in Stage III, more sensitive tests again reveal continuing deficits. Here, too, taste seems to be an important factor. Even though they are apparently eating normally, as little as 0.05% quinine hydrochloride w/v, mixed thoroughly in dry Purina powder, is often sufficient to cause rats with lateral hypothalamic lesions to refuse food completely to the point of extreme starvation. Normal rats readily accept food with up to 1.25% quinine, at which point they begin to decrease their intake and lose weight. When faced with the bitter taste of quinine, rats with lateral hypothalamic lesions, therefore, display a greater aversion to a negative taste in food. Palatability is again a powerful determinant of food intake in these animals. In the anorexic stage (Stage II), only highly palatable foods induced them to eat. And in the recovery stage, a slight negative taste prevents them from eating. This resembles the finicky eating observed in obese hyperphagic rats after more medial hypothalamic lesions (Teitelbaum, 1955). This finickiness implies either an impaired motivation for food (Miller, Bailey, & Stevenson, 1950; Teitelbaum, 1957) or an increased reactivity to its taste; presumably it is present in its most extreme form in the early stages of the lateral hypothalamic syndrome.

The recovery process continues even after the animal has started drinking water again. Some rats become more responsive to intraperitoneal injection of hypertonic saline and indeed three of seven rats tested eventually responded as well as a normal animal. Similarly in these same three animals it required as much as 0.1% quinine to prevent them from drinking whereas earlier they had been stopped by a 0.005 or 0.01% quinine solution. A gradual recovery can also be seen in the gross regulation of water intake, food intake, and weight. Immediately after entering the recovery stage, and for some time thereafter, some rats



	Stage I ADIPSIA, APHAGIA	Stage II ADIPSIA, ANOREXIA	Stage III ADIPSIA, DEHYDRATION-APHAGIA	Stage IV RECOVERY
EATS WET PALATABLE FOODS	NO	YES	YES	YES
REGULATES FOOD INTAKE & BODY WT. ON WET PALATABLE FOODS.	NO	NO	YES	YES
EATS DRY FOODS (IF HYDRATED)	NO	NO	YES	YES
DRINKS WATER. SURVIVES ON DRY FOOD AND WATER.	NO	NO	NO	YES

FIG. 3. Stages of recovery seen in the lateral hypothalamic syndrome. (The critical behavioral events which define the stages are listed on the left.)

with lateral hypothalamic lesions drink water in less than normal amounts, eat correspondingly less, and so maintain their weights at a subnormal level (210 to 220 grams). This hypodipsia may last for months, but usually it disappears. Such a rat begins to drink and eat more, and it regulates its weight at 270 to 280 grams, a more normal weight for its age.

The four stages of recovery from lateral hypothalamic lesions are summarized in Figure 3, which lists the critical behavioral events whose presence or absence defines those stages. First, there is a stage of "aphagia and adipsia" in which the animal completely refuses to eat or drink; second, "anorexia and adipsia" in which it eats, but not enough to maintain its weight, and does not drink; third, "adipsia and dehydration-aphagia," in which the animal fails to eat only because it does not drink; and finally the "recovery" stage in which the animal drinks water, eats dry food, and will survive, but still shows certain deficits in eating and drinking behavior. It must be emphasized that these stages in recovery will not be observed unless the animal is offered wet and palatable foods from the time of the operation. If only water and dry food are available, the rat will die unless it rapidly reaches Stage IV, at which point it accepts water and can, therefore, eat dry food.

ANATOMICAL ANALYSIS

The duration of these stages is variable and seems to depend on both the locus of the lesions and the amount of tissue damaged. The lesions in 13 animals were studied microscopically using thionin-stained celloidin sections through the diencephalon. Tissue in the lateral hypothalamic areas was destroyed bilaterally in all animals. Typically, the lesions were large, extending medially to the mammillothalamic tracts and the fornices (the fornix itself was only rarely destroyed) and laterally into the medial portions of the internal capsules and optic tracts. Damage was greatest at the level of the ventromedial nuclei. The lesions extended anteriorly to the level of the supra-optic nuclei and posteriorly into the pre-mammillary regions. Dorsally, the overlying subthalamus was always involved. In the majority of animals, the most ventromedial portions of the lateral hypothalamic areas were spared.

Figure 4 shows a frontal section showing lesions in the lateral hypothalamus at the level of the ventro-

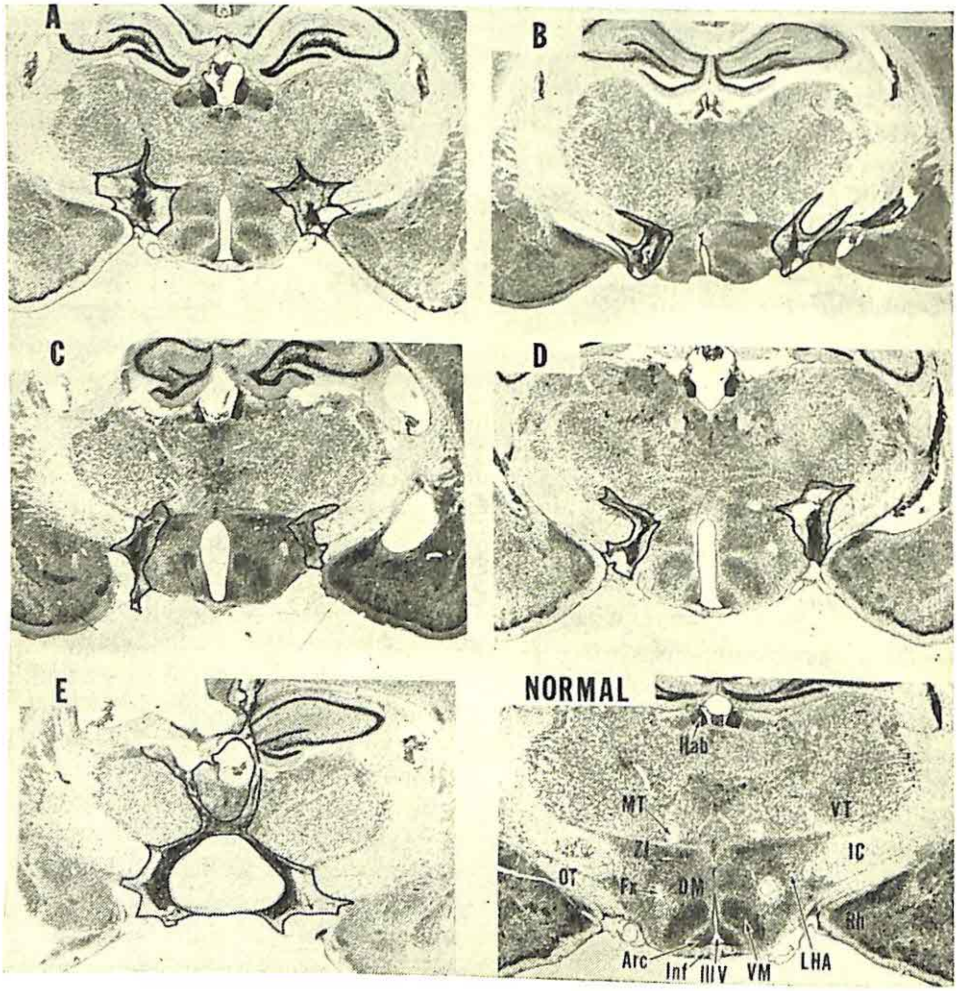


FIG. 4. Photomicrographs of thionine-stained sections through the tuberal region of the hypothalamus showing lesions in five animals with interesting variations of the lateral hypothalamic syndrome. (Damaged tissue is outlined in black. A: Large, symmetrical lesions producing permanent adipsia. Recovery graphed in Figure 1. B: Transient aphagia and anorexia, despite immediate postoperative hydration, followed by recovery of food and water intake. Animal 3 in Figure 2. C: Died on sixth postoperative day despite adequate hydration. Showed aphagia and anorexia. Animal 5 in Figure 2. D: Asymmetrical lesions producing transient adipsia without aphagia. E: Massive lesions involving both medial and lateral tissue producing hypothalamic hyperphagia and permanent adipsia.

Normal animal: Arc, arcuate nucleus; DM, dorsomedial nucleus; Fx, fornix; Hab, habenula; IC, internal capsule; Inf. infundibulum; LHA, lateral hypothalamic area; MT, mammillothalamic tract; OT, optic tract; Rh, rhinencephalon; VM, ventromedial nucleus; VT, ventral thalamus; ZI, zona incerta; III V, third ventricle.)

medial nuclei in five animals who exhibited interesting variations of the lateral hypothalamic syndrome. The extent of damage has been outlined heavily in black. With the very large lateral hypothalamic lesions of Section A, each stage in the recovery process lasted for weeks or months. This is the animal whose recovery is shown graphically in Figure 1. This rat showed prolonged aphagia and anorexia (38 days), and it never drank water again. Such an animal, because of the dehydration-aphagia that is present when it refuses to drink water, would seem to be permanently aphagic on a dry diet. It was only by inducing it to accept a palatable, wet diet that the stage of pure adipsia could be demonstrated. With smaller or asymmetrical lesions, as in Sections B, C, and D, the entire sequence passes very quickly, and the animal may begin the recovery not at Stage I, but at any of the later stages. These animals show aphagia and anorexia for only a few days or a week if offered a palatable liquid diet, and recover water intake within 1 or 2 months. Section B is from the brain of animal 3 in Figure 2. It refused to eat for 3 days despite automatic hydration, then nibbled wet food for 4 days, but remained adipsic for 1 month after operation. Section C shows the lesions of Animal 5 in Figure 2. It was aphagic and anorexic until it died even though hydrated. No attempt was made to keep this animal alive by tube-feeding, but since it had already begun to eat appreciable amounts of wet food, it probably would have regulated its caloric intake within a few more days. Section D shows lateral hypothalamic lesions which produced only adipsia, without any aphagia or anorexia. The animal accepted wet mash or liquid diet immediately and regulated its weight, but it refused to drink water and showed dehydration-aphagia on a dry diet. This effect closely resembles the earlier findings of adipsia without aphagia in the rat by Montemurro and Stevenson (1957) and in the dog by Andersson and McCann (1956). Finally, Section E, at a slightly more rostral level, shows a huge lesion that involves the entire hypothalamus except for the infundibulum, destroying both the ventromedial nuclei and the lateral hypothalamic areas, and extending laterally well into the internal capsules. This animal was aphagic and anorexic for only 4 days, then regulated its intake of wet food. It refused to drink water until it died 107 days later. When offered a wet diet it overate and became obese; but every time it was offered a dry diet and water, it refused to drink, stopped eating as a result of dehydration, and lost weight precipitously. This verifies the earlier findings of Williams and Teitelbaum (1959), who showed that hypothalamic hyperphagia can coexist with the lateral hypothalamic syndrome. It cannot be manifested in Stages I or II during which caloric regulation is not seen, but in Stage III, when caloric regulation appears, so does hypothalamic hyperphagia. But hydration is necessary. With a dry diet and water the only fluid available, the animal does not drink, and hypothalamic hyperphagia is masked by adipsia. In addition, such animals make it clear that damage to the ventromedial regions of the hypothalamus can produce hyperphagia even when the lateral hypothalamic areas are largely destroyed.

In general, the severity of the deficits was correlated with the size of the lesions. Transient deficits and rapid recovery followed small and asymmetrical lesions, whereas large lesions permitted only slow and incomplete recovery through stages of long duration. This suggested that the recovery of function was mediated by the undamaged tissue around the lesions. Additional lesions in the tissue immediately adjacent to the original damage were made in six recovered animals. In all of them, additional lesions were found to reinstate the syndrome. After appropriate lesions the reoperated animals became aphagic, anorexic, and adipsic all over again, and the recovery followed the same general pattern. The phenomenon was particularly clear in one animal. The recovery process was repeated four times. Each successive recovery, in this animal, was slower. The fourth recovery proceeded only to anorexia and was terminated by a fatal gastric hemorrhage. Other animals have been rendered permanently anorexic by additional lesions, in one for as long as 270 days. Therefore, the undamaged tissue surrounding the original lesion participates in the recovery. In addition, this result suggests that the experience gained in the initial recovery from lateral hypothalamic lesions did not protect the animal from the effects of additional damage.

Since the lateral hypothalamic syndrome in rats is the result of a complex interaction of deficits in both feeding and drinking, the lateral hypothalamus must contain tissues concerned with both functions. The structures involved could be cells within the lateral hypothalamus or fibers of passage from other areas. Morgane (1961a) reports a syndrome of aphagia and adipsia in the rat with lesions in the apex of the *globus pallidus,* completely outside the lateral hypothalamus. He points out that the major efferent pathways from the *globus pallidus* (*ansa* and *fasciculus lenticularis*) are usually interrupted by lateral hypothalamic lesions. Thus, extrahypothalamic structures participate in the regulation of food and water intake.

Whatever the nature of the neuroanatomical systems involved in feeding and drinking, some parts of them must overlap in the lateral hypothalamus. Is it possible to separate them? A chemical separation of the systems has been demonstrated in the rat by Grossman (1960). He showed that from the same site within the perifornical region of the hypothalamus, cholinergic drugs elicit drinking and adrenergic drugs produce feeding. This implies that the functions of each of the systems may be dependent upon different neurohumoral agents.

An anatomical separation may be possible because small, asymmetrical lesions produce adipsia without aphagia and pure effects have been produced by electrical stimulation. Drinking without feeding has been observed in the goat and rat (Andersson & McCann, 1955–56; Greer, 1955), as well as feeding without drinking (Larsson, 1954; Miller, 1957; Smith, 1956). Thus, it is quite possible that two systems which do not entirely overlap, one concerned with drinking and the other with feeding, merge in the lateral hypothalamus, and so destruction there yields the combination of deficits in feeding and drinking behavior that makes up the lateral hypothalamic syndrome.

Morgane (1961b) has suggested a further subdivision of the lateral hypothalamic feeding system. He finds that rats with lesions in what he has called the "far-lateral" region of the lateral hypothalamus (centered just medial to the internal capsule) lose weight more rapidly while suffering aphagia and adipsia and are less likely to recover than animals with lesions in the "mid-lateral" hypothalamus (centered just lateral to the fornix). According to Morgane, recovery is the rule following mid-lateral lesions. But recovery is the exception, if not an impossibility, after far-lateral lesions. He finds it difficult to maintain the far-lateral animal alive on a tube-feeding regime that is adequate for the mid-lateral animal. He suggests that the far-lateral region of the hypothalamus contains pallidofugal structures (*ansa* and *fasciculus lenticularis*) that are essential for the basic feeding reflexes. Lesions here should therefore produce permanent aphagia. In addition, the failure of the far-lateral animal to survive on tube-feedings leads Morgane to attibute its rapid decline to a form of "metabolic failure."

In our own experience, as described in this paper, many of the animals that have shown the typical stages of recovery were found to have extensive damage in the far-lateral hypothalamus (see Figure 4, especially Section A). Clearly these animals have very severe and prolonged deficits. In this we agree with Morgane. However, they do feed again, although they may remain adipsic. In the early stages of their recovery, they can be maintained on tube feeding if cared for assiduously. The more rapid decline in weight seen in similar animals and attributed to metabolic failure by Morgane can not, therefore, be due to an inability to utilize food but may be related to changes in activity, temperature regulation, or metabolic rate. Stevenson[6] has shown that lesions within the lateral hypothalamus may produce an increase in metabolic rate and has suggested that this increased energy expenditure may account for the rapid postoperative decline of the far-lateral animal. In addition, we must re-emphasize that lateral hypothalamic lesions produce deficits in *both* feeding and drinking. Prolonged deficits then may simply represent the summated effects of extensive destruction of both systems rather than selective involvement of "more basic" and functionally distinct portions of the feeding system alone.

SUMMARY AND CONCLUSIONS

Although we have made interpretations throughout this paper as new phenomena were described, there are several generalizations from this work that deserve emphasis. Each of them has changed our thinking about the lateral hypothalamic control of feeding and drinking.

We have confirmed the idea that there are tissues in the lateral hypothalamus that are concerned with both feeding and drinking. Aphagia is always accompanied by adipsia. We have never seen pure aphagia—i.e., absence of feeding where drinking is unimpaired. Any impairment in feeding then that follows these lesions can only be fully understood as the combined result of deficits in both regulations.

Recovery is the rule following these lesions, if the animals are maintained long enough. Barring accidents, or gastrointestinal pathology, it is possible to keep these animals alive indefinitely. Although we have seen prolonged deficits, we have never seen an animal with lateral hypothalamic lesions that did not show some recovery.

Feeding recovers before drinking. The quality of the diet is crucial for this observation. At first the animal will eat only wet and palatable foods. But if it is offered nothing except dry pellets and water, it will not eat until after it has begun to drink, and this relationship will be misunderstood. It appears to us, therefore, that lateral hypothalamic lesions impair feeding less severely than drinking.

---

[6] Stevenson, J. A. F. Personal communication, 1961.

Eating does not necessarily imply regulation. When the animals first begin to eat they are anorexic. They eat only highly palatable foods, and do not eat enough to regulate caloric intake or body weight. Their feeding seems to be evoked by the taste and smell of the food. They are drawn to it by appetite, not driven to it by hunger.

Hunger returns, often abruptly. A new control reappears, enabling the animal to regulate its caloric intake. The central feeding mechanism is now responsive to some variable associated with caloric need. If this variable were identified, we would know one control system that is sufficient for caloric regulation in the normal rat.

Even after the return of regulatory feeding, adipsia persists. It can mask caloric regulation if only dry food and water are available. The animal will not drink water and aphagia will be the result of dehydration. If the animal is induced to eat saccharin solutions, it hydrates itself involuntarily and then eats dry pellets. Thus, adipsia can occur without aphagia.

Adipsia can be permanent. Some animals never drink water again, though they live out their natural lives. In our experience, aphagia is always transient. We have never seen permanent aphagia after lateral hypothalamic lesions. This confirms our belief that the central neural tissue regulating thirst is more completely impaired by these lesions than is the tissue controlling feeding.

Drinking frequently reappears. Initially the animals drink only while they eat. Just as anorexia is feeding without response to starvation so prandial drinking is drinking without response to dehydration. Here again, ingestion does not necessarily imply regulation. Like hunger, thirst may eventually reappear. The central thirst mechanism is now responsive to dehydration.

Although regulation of food and water intake has recovered, motivation is still impaired. The recovered animal is greatly affected by the taste of both its food and water. Addition of small amounts of quinine deters it from eating and drinking. Wherever taste is excessively important, motivation is impaired.

Finally, in the lateral hypothalamic syndrome, the initial transition from aphagia to anorexia and from adipsia to prandial drinking seems to depend upon recovery of the influence of peripheral sensations (taste and smell of food, or dry mouth). The next transition involves the recovery of central neural controls permitting not only ingestion but ingestion with regulation. Anorexia gives way to hunger and prandial drinking to true thirst. This supports the commonly accepted view that feeding and drinking are both under multiple control (Brobeck, 1960; Stellar, 1954). Because all controls may operate simultaneously in the normal animal, it has been difficult to identify those essential for regulation. In the lateral hypothalamic syndrome, the reappearance of these controls is separated in time during recovery. They are not all operating simultaneously. Controls that operate in the later stages of recovery are absent in earlier stages. It should be possible to identify both the controls that are present and those that are absent during any particular stage by subjecting the animal in that stage to the variables known to influence feeding and drinking in the normal animal. If we find controls that are absent before regulation, and which reappear with it, we may then be able to specify the controls necessary for regulation.

REFERENCES

ADOLPH, E. F. Ontogeny of physiological regulations in the rat. *Quart. Rev. Biol.*, 1957, **32**, 89–137.

ADOLPH, E. F., BARKER, J. P., & HOY, P. A. Multiple factors in thirst. *Amer. J. Physiol.*, 1954, **178**, 538–562.

ANAND, B. K., & BROBECK, J. R. Hypothalamic control of food intake. *Yale J. Biol. Med.*, 1951, **24**, 123–140.

ANAND, B. K., DUA, S., & SCHOENBERG, K. Hypothalamic control of food intake in cats and monkeys. *J. Physiol.*, 1955, **127**, 143–152.

ANDERSSON, B., & MCCANN, S. M. Drinking, antidiuresis and milk ejection from electrical stimulation within the hypothalamus of the goat. *Acta physiol. Scand.*, 1955–56, **35**, 191–201.

ANDERSSON, B., & MCCANN, S. M. The effect of hypothalamic lesions on the water intake of the dog. *Acta physiol. Scand.*, 1956, **35**, 312–320.

BROBECK, J. R. Regulation of feeding and drinking. In J. Field (Ed.), *Handbook of physiology. Sect. 2, Vol. 2.* Baltimore: Williams & Wilkins, 1960.

BROBECK, J. R., TEPPERMAN, J., & LONG, C. N. H. Experimental hypothalamic hyperphagia in the albino rat. *Yale J. Biol. Med.*, 1943, **15**, 831–853.

BRÜGGER, M. Fresstrieb als hypothalamisches Symptom. *Helv. physiol. pharmacol. Acta*, 1943, **1**, 183–198.

CARPER, J. W., & POLLIARD, F. A comparison of the intake of glucose and saccharin solutions under conditions of caloric need. *Amer. J. Psychol.*, 1953, **66**, 479–482.

EPSTEIN, A. N. Reciprocal changes in feeding behavior produced by intrahypothalamic chemical injections. *Amer. J. Physiol.*, 1960, **199**, 969–974. (a)

EPSTEIN, A. N. Water intake without the act of drinking. *Science*, 1960, **131**, 497–498. (b)

EPSTEIN, A. N., & TEITELBAUM, P. Regulation of food intake in the absence of taste, smell, and other oro-pharyngeal sensations. *J. comp. physiol. Psychol.*, in press.

FRENCH, J. D., PORTER, R. W., VON AMERONGEN, F. K., & RANEY, R. B. Gastrointestinal hemorrhage and ulceration associated with intracranial lesions. *Surgery*, 1952, **32**, 395–407.

GREER, M. A. Suggestive evidence of a primary "drinking center" in hypothalamus of the rat. *Proc. Soc. Exp. Biol. Med.*, 1955, **89**, 59–62.

GROSSMAN, S. P. Eating or drinking elicited by direct adrenergic or cholinergic stimulation of hypothalamus. *Science*, 1960, **132**, 301–302.

LARSSON, S. On the hypothalamic organization of the nervous mechanism regulating food intake. *Acta physiol. Scand.*, 1954, **32**, Suppl. No. 115, 7–40.

MILLER, N. E. Experiments on motivation. *Science*, 1957, **126**, 1271–1278.

MILLER, N. E., BAILEY, C. J., & STEVENSON, J. A. F. Decreased "hunger" but increased food intake resulting from hypothalamic lesions. *Science*, 1950, **112**, 256–259.

MONTEMURRO, D. G., & STEVENSON, J. A. F. Adipsia produced by hypothalamic lesions in the rat. *Canad. J. Biochem. Physiol.*, 1957, **35**, 31–37.

MORGANE, P. J. Alterations in feeding and drinking behavior of rats with lesions in globi pallidi. *Amer. J. Physiol.*, 1961, **201**, 420–428. (a)

MORGANE, P. J. Medial forebrain bundle and "feeding centers" of the hypothalamus. *J. comp. Neurol.*, 1961, **117**, 1–26. (b)

MORRISON, S. D., & MAYER, J. Adipsia and aphagia in rats after lateral subthalamic lesions. *Amer. J. Physiol.*, 1957, **191**, 248–254.

SMITH, O. A. Stimulation of lateral and medial hypothalamus and food intake in the rat. *Anat. Rec.*, 1956, **124**, 363–364.

STELLAR, E. The physiology of motivation. *Psychol. Rev.*, 1954, **61**, 5–22.

STELLAR, E., & HILL, J. H. The rat's rate of drinking as a function of water deprivation. *J. comp. physiol. Psychol.*, 1952, **45**, 96–102.

STELLAR, E., & KRAUSE, N. P. New stereotaxic instrument for use with the rat. *Science*, 1954, **120**, 664–666.

STEVENSON, J. A. F., & RIXON, R. H. Environmental temperature and deprivation of food and water on the spontaneous activity of rats. *Yale J. Biol. Med.*, 1957, **29**, 575–584.

TEITELBAUM, P. Sensory control of hypothalamic hyperphagia. *J. comp. physiol. Psychol.*, 1955, **48**, 156–163.

TEITELBAUM, P. Random and food-directed activity in hyperphagic and normal rats.

J. comp. physiol. Psychol., 1957, **50**, 486–490.

TEITELBAUM, P., & STELLAR, E. Recovery from the failure to eat produced by hypothalamic lesions. *Science*, 1954, **120**, 894–895.

WHEATLEY, M. D. The hypothalamus and affective behavior in cats. *Arch. Neurol. Psychiat., Chicago*, 1944, **52**, 296–316.

WILLIAMS, D. R., & TEITELBAUM, P. Some observations on the starvation resulting from lateral hypothalamic lesions. *J. comp. physiol. Psychol.*, 1959, **52**, 458–465.

WYRWICKA, W., & DOBRZECKA, C. Relationship between feeding and satiation centers of the hypothalamus. *Science*, 1960, **132**, 805–806.

(Early publication received August 14, 1961)

Psychological Review
1962, Vol. 69, No. 2, 91–108

A STATISTICAL THEORY OF DYNAMIC CONTOUR PERCEPTION [1]

W. M. Smith
Dartmouth College

and

W. L. Gulick
University of Delaware

Bartley's (1941) statement made almost twenty years ago, ". . . that the matter of contour formation is one of the central phenomena of sensory psychology and nerve physiology" (p. 334), certainly is as true today as it was when he wrote it. Indeed, the way in which borders are perceived is a fundamental problem in visual science.

It is very clear that the classical idea of the retinal mosaic as a system of insulated receptors cannot meet adequately various facts of visual acuity, brightness discrimination, and contour definition. Visual acuity is greater than the classical notion implies, and contour perception is not dependent upon two parallel arrays of receptors, one array stimulated, the other not. Inadequate consideration of postretinal structures has been the major failing of what is now considered to be the classical view of visual acuity.

The work of Marshall and Talbot (1942) is a formulation of a theory of visual acuity and contour resolution in terms of the structure and the function of retinal *and postretinal* neural elements. Drawing upon recent findings and concepts of neurophysiology, they view such basic processes of acuity and contour perception as dependent upon the interrelated activity of *populations* of receptors and neural elements. Appropriately, their formulation now is referred to as a statistical theory of visual function.

Although we shall not present a review of the Marshall and Talbot work, for expository purposes later it will be helpful to mention some general features of their theory here. Their analysis rests heavily upon the neurological concepts of reciprocal overlap, lateral and vertical summation in the visual system, physiological nystagmus, and the nature of the neural recovery cycle. Especially important is the fact that their theory places the locus of the critical aspects of contour perception in the visual cortex rather than at the retinal level.

In their original work Marshall and Talbot were concerned primarily with the general question of sensory acuity, but in more recent years other investigations have extended the statistical theory to other aspects of vision. Osgood and Heyer (1952), for example, have applied it to the analysis and interpretation of figural aftereffects and have claimed the statistical theory to be superior to the field conception offered by Köhler and Wallach (1944). Regarding figural aftereffects, however, both the field and the statistical interpretations have had their critics (Deutch, 1956; Smith, 1952). More recently, Day (1956, 1957) has used statistical theory in the interpretation of phenomenological reports and experimental data concerning form perception. In particular, he has employed statistical theory in considering the question of the emergence of form in the peripheral retina and in interpreting data from studies of area and area-perimeter ratios in form perception. Finally, Osgood (1953, pp. 245–47) has presented a brief account of how statistical theory can be applied to the problem of apparent movement.

---

[1] This work has been supported by the National Science Foundation.

It is with the perception of movement that the present paper is concerned. We shall attempt to extend statistical theory to a treatment of perceived movement. First, we shall describe briefly the basic experimental situation. Then we shall summarize the results from which the theoretical treatment was developed. Finally, we shall present an account of our modification and extension of statistical theory with supporting evidence.

DYNAMIC CONTOUR PERCEPTION

For present purposes we define "dynamic contour perception" as the perception of the borders of moving stimuli. The definition could be modified to include instances wherein the observer moves rather than the stimulus. We shall not consider this modification here.

In two previous papers (Smith & Gulick, 1956, 1957) data have shown that the perception of the borders of moving stimuli is influenced not only by the velocity of the stimulus, but also by the length of time the stimulus is observed in a fixed position prior to movement. The results of these experiments indicated that the border of a small square could not be seen clearly at velocities exceeding about 13 degrees of visual angle per second. However, when the stimulus was presented in a fixed position prior to movement, the maximum velocity at which the edges could be perceived was increased by a factor or two or more in most instances. A systematic study of the combinations of velocity and initial exposures needed to maintain contour revealed a linear relationship up to exposures of about 350 milliseconds and velocities of about 30 degrees of visual angle per second. For every increase in velocity that produced loss of contour, an increment in the duration of the initial exposure could be utilized to regain it. In more recent unpublished studies in which the duration of exposure of the initial stationary phase of the stimulus has been extended to 1 second, we find that the ability to see an edge with increasing velocities of movement is related curvilinearly to the duration of the initial stationary phase. At velocities exceeding about 40 degrees per second it appears that the fixed phase of presentation, regardless of its duration, will not enable one to see a moving edge clearly. In an earlier paper (Smith & Gulick, 1957) we offered a preliminary statement of statistical theory as applied to the combined effects of the initial exposure of the stimulus in the fixed position and the movement phase. In the present paper we wish to expand this preliminary theoretical treatment.

HISTORICAL BACKGROUND OF THE MODEL

When an observer's total visual field consists of a light-dark pattern divided vertically, there is no apparent gradient of gray where light meets dark. The discontinuity is clearly defined. How the visual system is able to "construct" an edge from such a distribution of light energy striking visual receptors is not well understood. However, it is known that the abrupt gradient of reflected light which exists at the border of a stimulus object is not present once the pattern of light reaches the retina (Fry, 1955). The abrupt gradient becomes less abrupt so that the intensity of light comprising the geometric image of a border at the retina approximates the integral of a normal distribution. It is this pattern of light energy which courses back and forth

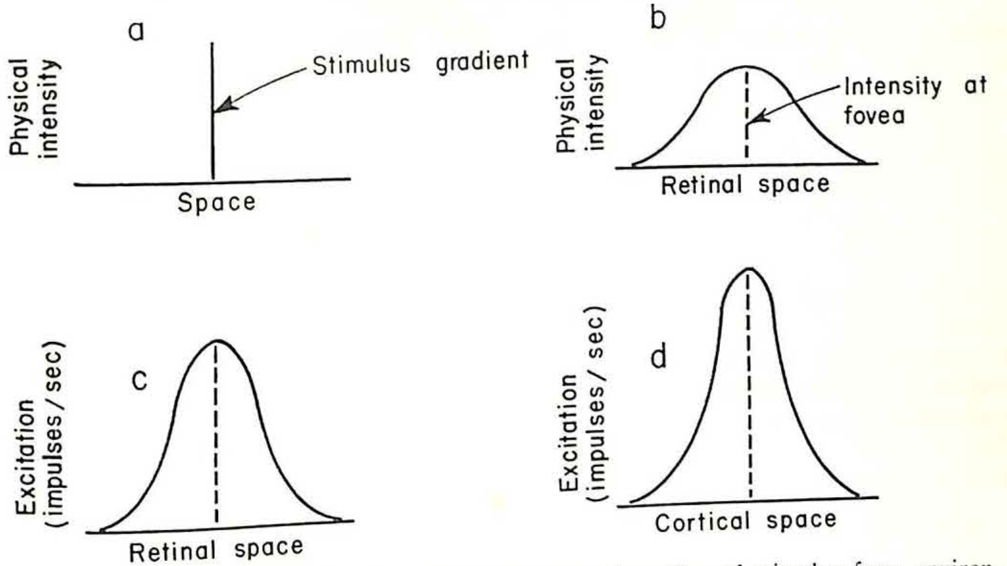


FIG. 1. Hypothetical distributions illustrating transformation of stimulus from environment to cortical space. (The platykurtic distribution of light at the fovea, b, is transformed to a more leptokurtic distribution of excitation within the afferent system, c. At the cortical level, d, the distribution of excitation is peaked further and is the direct correlate of perceived contour.)

across a population of receptors as the eye undergoes nystagmic movements.

If, instead of a single border, an observer is shown a pattern of two parallel borders, such as a bright, thin vertical line (near threshold value), the distribution of light of this pattern on the retina approximates a normal curve. As the angular width of such a pattern increased, the symmetrical distribution at the retina would be changed progressively, becoming flattened on top. Each border, however, would be represented on the retina by a distribution of light similar to that for the discontinuity of the vertically divided field mentioned previously.

It is the central proposition of statistical theory that the *cortical* representation of a perceived edge is a symmetrical distribution of excitation in Area 17. The leptokurtic distribution of excitation is correlated with the abrupt change in intensity at the stimulus boundary, but only by virtue of a very complicated series of events and mechanisms intervening between cornea and cortex. The abrupt change of energy level in the stimulus is not reproduced at the retina because of distortion introduced by passage through ocular media. Slight irregularities of the surface of the eye, errors of refraction and of aberration, and entopic stray light all tend to reduce the abruptness of the stimulus gradient as it is projected on the retina.

In Figure 1 we have attempted to show schematically the transformation from stimulus gradient to cortical distribution. A line in the visual field presents an abrupt gradient of energy. At the retina the distribution of energy assumes the platykurtic form as shown in Figure 1b. The relative flatness of this distribution is due to the scattering of light by the distorting factors noted previously, plus the effects of physiological nystagmus. This distribution of light represents the proxinal stimulus which triggers the primary receptors. If one plots the rate of neural firing for different retinal loci, then a distribution of excitation results. This hypothetical distribution is shown in Figure 1c. The peaking of this distribution is produced by the combined effects of numerous mechanisms. First, neural elements located near the center of the distribution in Figure 1b are stimulated by the most intense light. Accordingly, the number of fibers stimulated as well as the average rate of firing of these neurons is greater than in those neural elements located near the edges of the distribution. Second, as a result of physiological nystagmus, centrally located receptors are stimulated more efficiently because the rate of application of the light energy is maximized in those receptors located near the maximum slope of the distribution of light (Rushton, 1935). If we assume the primary "carrier" of contour activity to be the on-off fibers, then these fibers serving receptors centrally located will have a higher rate of firing than those serving receptors found at the periphery of the population of receptors involved. Third, because of the faster rate of application of stimulation just noted, the amount of neural accommodation in these centermost fibers will be minimized. Neural accommodation, according to Nernst (1908), is a counter-process of some kind in the neuron which cuts short the effect of a stimulus. As the rate of application of a stimulus is increased, the amount of accommodation is decreased (Monnier, 1934). Whether neural accommodation actually operates in the vertebrate retina remains to be demonstrated conclusively, but there is some evidence to suggest that it does (Granit & Skoglund, 1943). The result of the increased rate of application of the stimulus, and hence decreased accommodation, produces an additional relative increase in excitation rate for the neurons associated with the center of activity in the retina (Skoglund, 1942).

There are, in addition, possible bases for peaking the excitation through the action of retinal receptive fields (Barlow, 1953a, 1953b; Hartline, 1940; Kuffler, 1953). A receptive field of a particular ganglion cell typically is organized with antagonistic discharge patterns of the center and the periphery, that is, the center may be predominantly "on" and the periphery of the field "off," or vice versa. The intermediate area of the field will be "on–off." Thus, any displacement of a border across such a field as might be produced by slight movements of the eye will result in continuous excitation of the ganglion cell. If the field is an 'off-center" type and the gradient of stimulation passes from center to periphery then there will be bursts of excitation on leaving center, entering and leaving the intermediate area, and upon entering the peripheral part of the field. Any kind of scanning back and forth across this field obviously will maximize excitation of the ganglion cell associated with the receptive field. If to these complex events occurring at the retinal level we add the effects of reciprocal overlap occurring during projection, the distribution of excitation in the visual cortex is peaked further and we have the leptokurtic distribution as shown in Figure 1d.

The phenomenon. In our early experiments, a stimulus (A) was observed moving horizontally through distance (D), from left to right. Its velocity (V) was increased until the edges of A no longer could be seen clearly. At this velocity, if A were exposed briefly in a fixed position prior to movement (this duration we called

t_1), *A* was seen with sharp edges. This effect was systematically investigated for velocities from 0 to approximately 30 degrees of visual angle per second. When values of *V* were plotted against values of t, a linear relationship was obtained. In order to maintain the contour of a moving stimulus, each increase in *V* of 1 degree visual angle per second required a corresponding increase in t of approximately 25 milliseconds.

The addition of the prior stationary phase to movement definitely facilitated the perception of the edges of the moving stimulus. Stemming from a statistical treatment of apparent movement (Osgood, 1953, pp. 245–47), wherein two separate distributions of excitation in the cortex sum into one distribution whose peak is displaced in cortical space, a tentative explanation of our contour data was offered. The hypothesis was that excitation arising from the moving stimulus summed with the decaying activity around c_1 (the cortical locus of maximum activity produced by the fixed stimulus) to generate a single peaked distribution moving through cortical space toward locus c_2 (the cortical locus of maximum activity produced by the moving stimulus at the termination of movement).

THE MODEL

It is now appropriate to state some assumptions upon which our statistical concept of dynamic contour is based. Some of these assumptions follow closely those made by other statistical theorists; others are especially germane to the question of dynamic contour.

1. On-off and off receptors are the principal retinal mediators of contour perception. The excitation ensuing from the stimulation of these two types of receptors becomes the cortical representation of a border. Because of their response characteristics it is assumed that on fibers play a minor role.

2. Each of an infinite number of points along a border of a stimulus (hairline) is represented in the visual cortex by a normal or near-normal distribution of excitation (rate of firing of neural elements extended over cortical space).

3. The continuum of points along the border of the stimulus is represented in the cortex by a continuum of points of maximal excitation, the longitudinal extent of which is proportional to the length of the border.

4. The mean slope of the cortical excitation distribution, rather than the absolute amount of excitation, defines the clarity of an edge. The sharpness of the edge is determined directly by the relationship between the amount of excitation at its maximum and the amount of dispersion of excitation through cortical space. It is assumed that the mean slope of the distribution of excitation increases through time in a negatively accelerated fashion. Only when the slope equals or exceeds some critical value range is contour perceived. This implies some finite period of time as a necessary antecedent to contour perception.

The slope of the excitation distribution in the cortex at any instant in time is determined by the intensity of light, contrast, and by the state of the eye and proximal neural elements.

5. Overlapping distributions of excitation in the cortex summate.

6. The mean slope of the distribution of excitation also is influenced by velocity of movement: slope is inversely related to velocity.

7. After cessation of stimulation arising from a fixed or moving border, the mean slope of the excitation distribution decreases through time in a

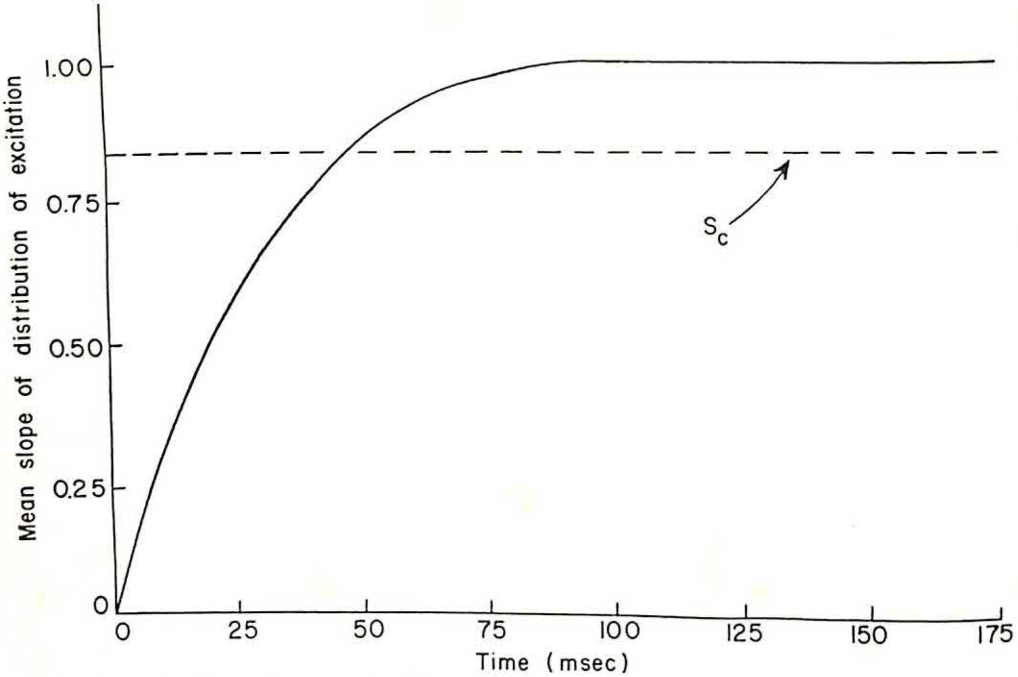


FIG. 2. Hypothetical curve depicting the increase in the mean slope of the excitation distribution through time. (S_c—critical slope—signifies the value necessary for the perception of contour. The slope of the function in the figure would change with various stimulus conditions. In this example, contour is perceived about 50 milliseconds after initiation of stimulation.)

negatively accelerated fashion. The mean slope at any given instant in time during decay is influenced by the intensity of the light: the higher the intensity, the steeper the slope.

When light stimulates the retina, the level of cortical excitation related to it grows in a negatively accelerated manner. Moreover, and of more importance, the excitation distribution representing a contour increases in mean slope through time in a negatively accelerated fashion. This increase in slope is shown in Figure 2.

This hypothetical function indicates that in approximately 50 milliseconds the slope (S) reaches a value necessary for sharp contour perception (S_c). Although this temporal value is to some extent arbitrary, it illustrates here that the appearance of sharp edges in the visual field takes a finite period of time to develop. There is no single instant when a sharp edge appears. Rather, the transition from "fuzziness" to a clearly defined edge is a continuous matter. The value of 50 milliseconds in Figure 2 represents a general threshold estimate for the appearance of a well-defined edge. Numerous factors can influence the rapidity with which S_c is attained, among which are intensity level, degree of contrast, and level of retinal adaptation. Regardless of specific conditions of observation, however, the delay in reaching S_c after the onset of stimulation is due to the temporally dispersing factors of synaptic delay, reverberation, and horizontal spread of neural activity as it proceeds from the receptors to the visual cortex.

When a border is removed from the visual field, its related excitation distribution decays with negative acceleration. Furthermore, the slope of this distribution also decreases and in the same general manner (Assumption 7). Figure 3 represents a hypothetical decay function showing the decrease in mean slope of the excitation distribution as a function of time. Note that this decay function is influenced to an important degree by the duration of the stimulus and its intensity. The prolonged decline in the mean slope is due to the continued excitation as a result of the reluctance of the on-off and off fibers in the optic nerve to cease activity. The nature of this decay is based primarily upon evidence obtained by Hartline (1938) of residual activity persisting for many seconds, although gradually subsiding. Because the slope during decay exceeds S_c for some brief but finite period of time after the cessation of stimulation, an edge can be perceived during this brief interval.

The hypothetical functions depicted in Figures 2 and 3 represent the increase in mean slope as a function of time produced by an initial stationary stimulus (Figure 2) and the decrease in mean slope as a function of time produced by the cessation of the same stimulus (Figure 3). When a *moving* stimulus is presented to the eye, both the mean slope of excitation and the level of excitation increase in the same general manner as that already discussed for stationary stimuli. In Figure 4 are shown three hypothetical functions describing the mean slope, one for each of three velocities of movement. Assuming other conditions constant, they illustrate that the time required to attain S_c is determined by velocity of movement. In Assumption 6 it was stated that the mean slope of the distribution of excitation produced by a stimulus in motion was affected by the velocity of

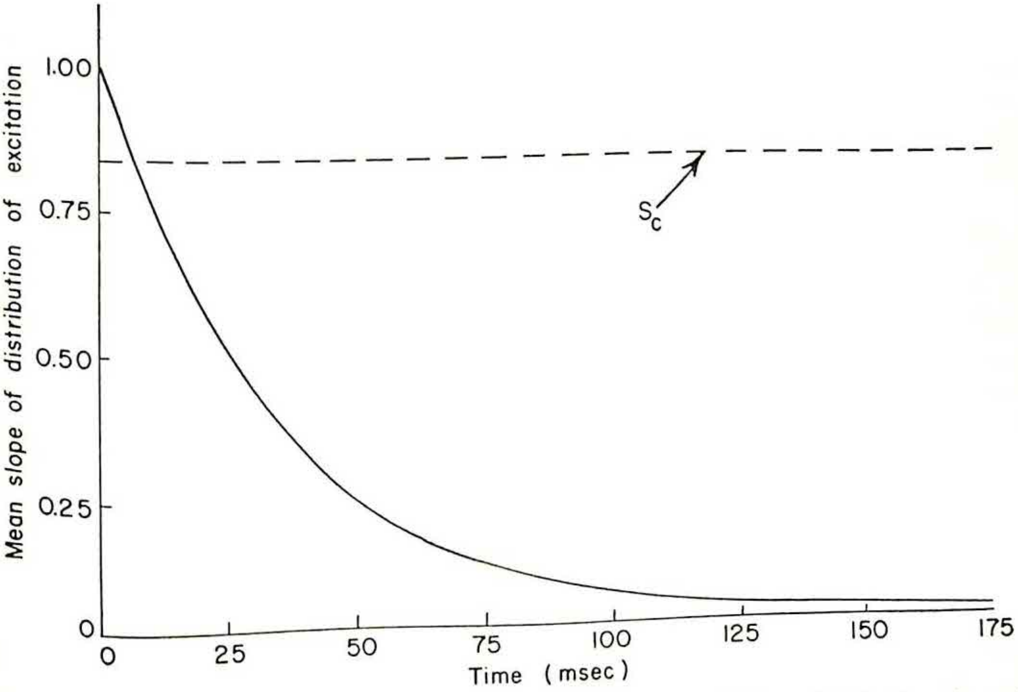


FIG. 3. Hypothetical curve of decay in mean slope of excitation distribution through time. (In this example, contour is perceived for a few milliseconds after cessation of stimulation until the slope of the excitation distribution falls below S_c.)

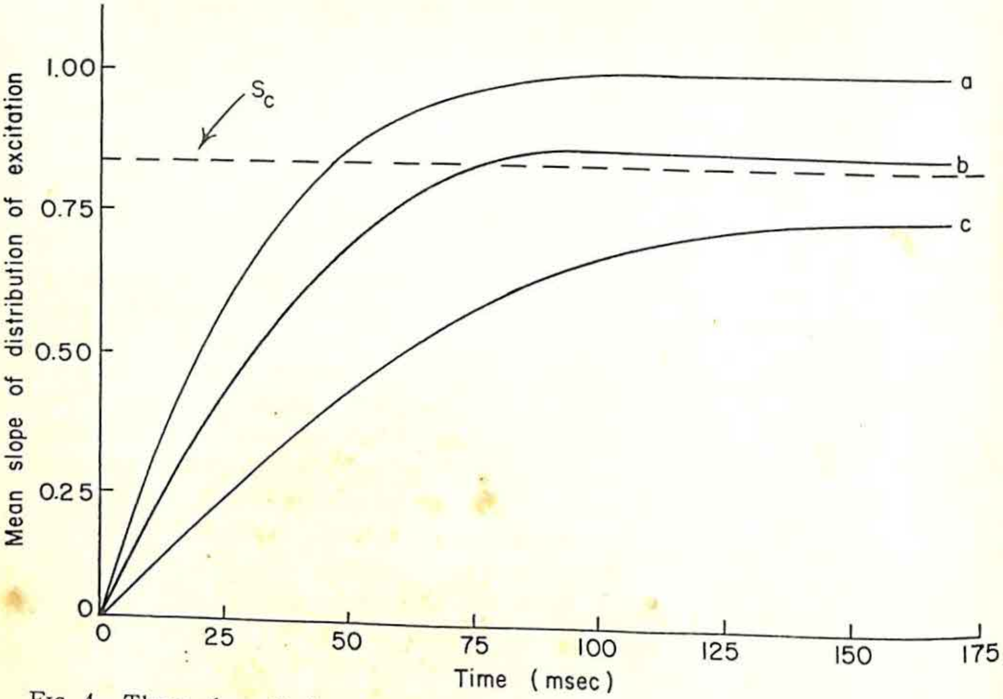


FIG. 4. Three theoretical curves showing how the increase in mean slope of the excitation distribution changes over time for different velocities of movement. (As velocity of movement increases, more time is required for the mean slope to reach S_c, the level necessary for contour perception. At high velocities S_c may not be attained—Curve c.)

movement. At low velocities ($V <$ 10°/sec) the time required to attain S_c is approximately equal to that required to attain it with a stationary stimulus (Curve a). However, at higher velocities ($V > 10°$/sec) more time is required before S_c is attained (Curve b). Indeed, if V is sufficiently high, S_c will never be reached (Curve c). As illustrated with these hypothetical functions, when velocities are very high (Curve c) the mean slope of the excitation distribution never reaches the critical value, S_c. Therefore, clearly defined edges cannot be perceived. It is under these conditions that the presence of the initial stationary phase of the stimulation sequence becomes useful.

What is now required is to elucidate the manner in which the excitation arising from the stationary stimulus sums with the excitation produced by the subsequent moving stimulus so as to maintain jointly a value of S_c which is sufficient for contour perception. It is our hypothesis that the maintenance of contour during stimulus movements as a result of the presentation of the initial stationary phase is achieved by the summation of the distributions of excitation arising from the total stimulation sequence. Excitation resulting from the moving stimulus is added to the residual remaining after the withdrawal of the stationary stimulus, and their sum under certain conditions can be sufficient to increase the mean slope of the combined excitation distribution to equal or surpass S_c at a time much earlier than would otherwise have obtained had S_c been attained solely on the basis of excitation from movement.

Even under conditions of static contour perception a brief interval is re-

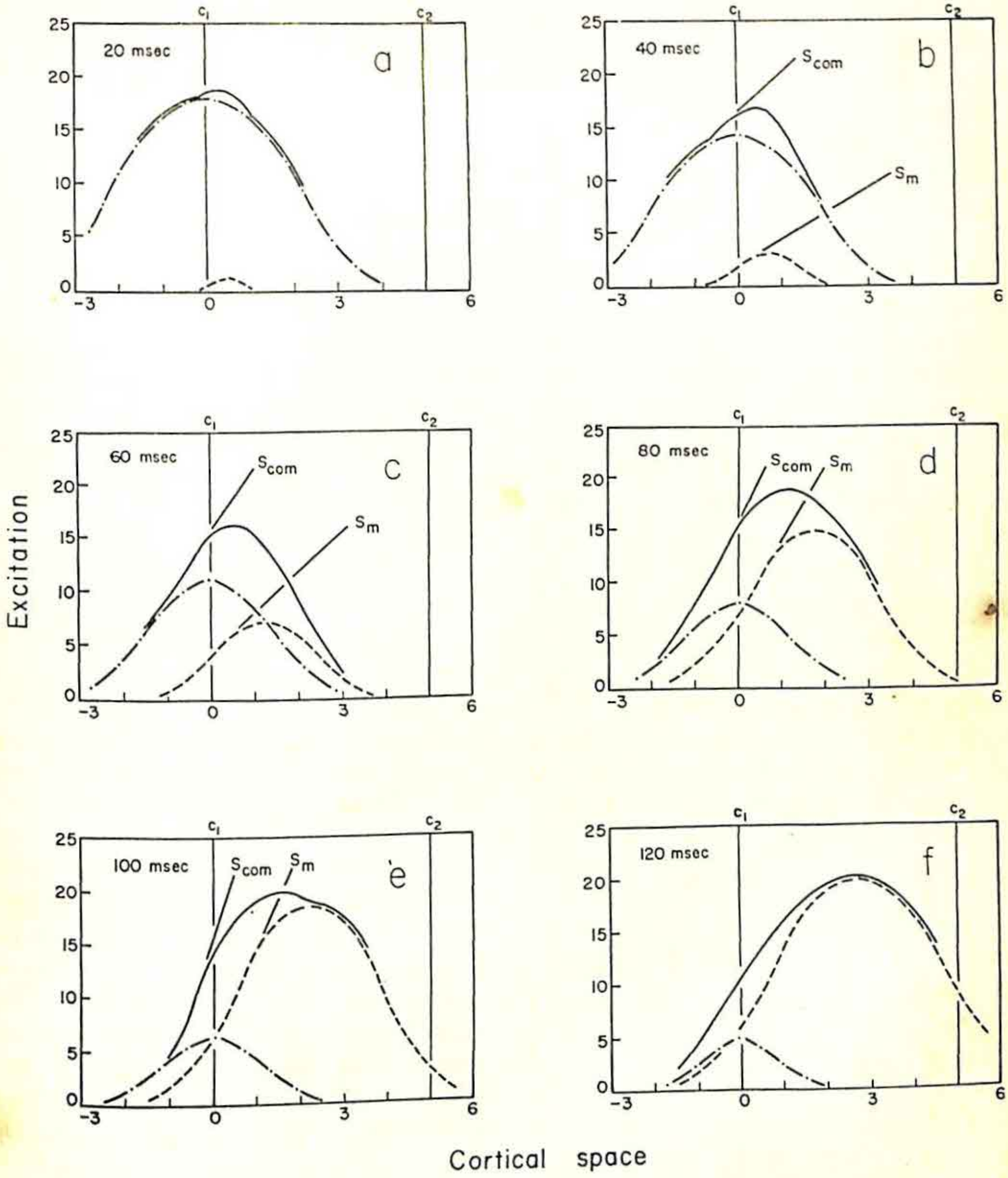


FIG. 5. Theoretical curves showing how the combined distributions of excitation for the moving contour (regular broken lines) and residual and decaying excitation from the initial fixed exposure of the stimulus (dot-dash lines) combine (solid curves) over time to produce contour perception. (As time increases from the moment of the initiation of movement to 20 milliseconds, a; 40 milliseconds, b; 60 milliseconds, c; etc., the level of excitation related to the moving border increases as it moves through cortical space. The level of excitation of the combined distribution also increases through time. Here it may be seen that the slope of the distribution representing the moving stimulus, S_m, increases through time. The slope of the combined distribution, S_{com}, increases through time at a faster rate, and thereby hastens the attainment of the critical slope, S_c, for contour resolution. The symbol C_1 represents the cortical location representing the peak of excitation from the stationary phase, and C_2 the cortical location of the peak of excitation from the moving phase at the end of movement. The abscissa represents cortical space in arbitrary units.)

quired before S_e is reached because of the temporally dispersing factors already mentioned. This interval probably increases slightly when a stimulus moves slowly; but when velocity exceeds 10 degrees or so per second, the interval increases markedly. Within the range of velocities for which the initial fixed stimulus is required in order to maintain contour, larger and larger residuals of excitation must be added to the excitation produced by movement if contour is to be present. The result of this summation is to increase the slope of the combined excitation distribution, and hence hasten the attainment of S_e.

This summation is illustrated for increasing temporal intervals in Figure 5. Note in Figure 5 that the excitation distribution for the fixed stimulus (dot-dash line) decays through time, thereby producing a distribution of progressively less slope. Note also that the slope of the excitation distribution representing the moving stimulus (dashed line) increases through time, as does the combined distribution (solid line). The most important point illustrated by the set of curves, however, is that the mean slope of the combined distribution increases at a faster rate than does the mean slope of the excitation distribution for movement. This means that the perception of the edge of the moving stimulus is facilitated and can be seen at an earlier point in time than would be the case had there not been the contribution to excitation by the decay of the stimulation from the fixed stimulus.

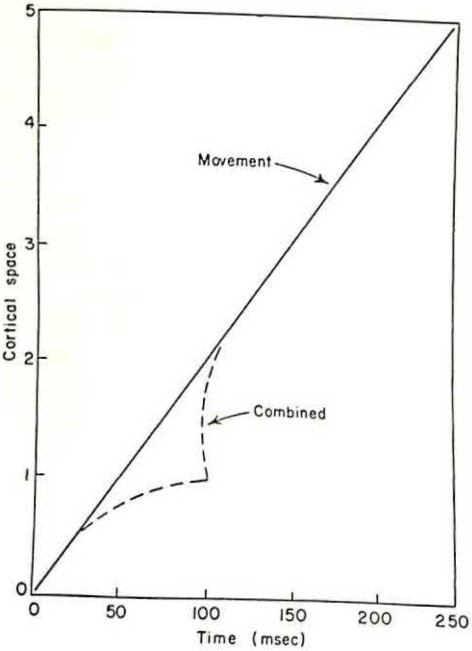


FIG. 6. When the velocity of movement of a stimulus is constant, the rate of displacement of the peak of excitation in the cortex is also constant (solid line). However, when the decaying activity from the fixed stimulus sums with the growth of activity produced by the moving stimulus, the peak of the combined distribution is not displaced through cortical space at a constant rate (dashed line).

Figure 5 demonstrates another important aspect of this summation process. According to the Marshall and Talbot (1942) view, borders of a stimulus are represented by the peaks of distributions standing for points along the borders or in their terms, a ridge of excitation is produced by an infinite number of points along the edge of a stimulus. For static contour perception the location of the stimulus edge in the visual field and its cortical location bear a rather straightforward relationship. In dynamic contour perception, however, this relationship becomes more complex. If we assume the eye to be fixed, when the velocity of movement of the stimulus object is constant, so is the velocity of the moving peak of excitation as it travels through cortical space. In dynamic contour perception based upon the summation of excitation distributions the peak of total excitation does not travel through cortical space at a constant velocity. In Figure 5d the peak of the combined distribution can be seen to lag behind the peak produced by the moving border alone (by about one unit of cortical space). The difference in the rates of displacement for the peak of the distribution of excitation produced by the moving border and the peak of the summed distribution of excitation are presented in Figure 6. This figure illustrates that the rate of displacement of the peak of excitation for the moving stimulus alone is constant; hence the straight line. However, the rate of displacement of the peak of excitation for the combined distribution is not constant, but first negatively accelerated and then positively accelerated. It is hypothesized that this change in rate of displacement of the peak of the distribution of the summed excitations is responsible for a peculiar effect reported by our subjects; namely, that the velocity of the moving edge did not appear to be constant. This account of perceived variable velocity with constant velocity of movement we believe to be one important piece of supporting evidence for the theory presented here.

To summarize: For low velocities of movement S_c is attained with sufficient speed to allow virtually instantaneous contour perception. With high velocities S_c is never attained during the time interval required to complete the stimulus movement sequence. For the intermediate range of velocities (8–30°/sec), S_c is attained rapidly enough to permit contour perception only if the decaying neural activity produced by the fixed stimulus sums with the growth of neural activity produced by the moving stimulus.

SUPPORTING EVIDENCE

Lack of techniques for specific measures of the integrative activity of the visual system makes difficult any direct test of statistical theory as expressed here. However, indirect evidence of a supporting nature can be obtained from various perceptual experiments. We have mentioned previously the evidence concerning the perceived variable velocity of a constant velocity stimulus. To other supporting evidence we now turn.

Contrast. We have said that contour perception cannot occur unless the slope of the cortical distribution representing contour reaches some minimal value (S_c). One factor which should have a direct and important bearing on the activity of on-off fibers, and hence the peaking of the related cortical distribution of excitation, is the intensity difference representing a border. It was hypothesized, therefore, that under conditions of high contrast between a stimulus object and its surround, the perception of a moving edge could occur at velocities of movement higher than those which would be expected under low contrast. Because high contrast would decrease the time required to obtain S_c, the importance of the summative effects of decaying neural activity following the presentation of the initial stationary stimulus would be minimized. That is, a stimulus of high contrast at a given velocity should require less contribution from the decaying effects of the initial stationary stimulus than a stimulus of the same velocity and lower contrast. The major contribution of the initial stationary stimulus is to provide, after its cessation, neural excitation which may sum with that produced by the moving stimulus. This summation effect simply hastens the attainment of S_c during movement (see Figure 5).

In terms of our operations the result of an increase in contrast would be to allow the perception of contour of the moving stimulus at a higher

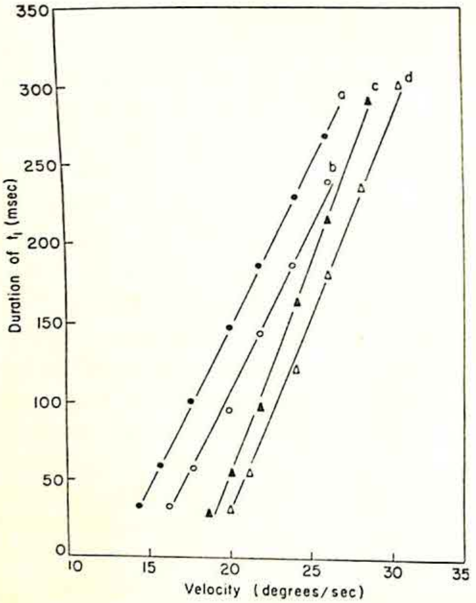


FIG. 7. The effects of contrast and illumination level on contour perception of moving stimuli. (Data of five subjects. Brightness values in footlamberts of the background and the stimulus: d, 4.2–1.2; c, 4.2–1.5; b, .036–.007; a, .036–.015. The stimulus was always darker than the background. Stimulus size equaled 0.5 degrees, and the extent of movement was 5.0 degrees of visual angle. Each curve in the figure specifies how long the stimulus had to appear in a fixed position prior to movement, t_1, in order for contour to be perceived during movement of a particular velocity.)

velocity than under conditions of low contrast, even though the duration of the fixed stimulus remained constant. Conversely, for a given velocity of movement, a reduced exposure interval of the initial fixed stimulus is adequate to maintain contour perception during movement whenever contrast is increased.

Data on the effects of contrast are presented in Figure 7. For each of five subjects, velocity of movement was increased from 8 degrees per second in 2-degree steps until contour was lost (about 13°/sec). Thereafter, the value of t_1 required to see contour was determined for each increase in V. In Figure 7 mean values of t_1 are plotted against V for each of two contrast conditions at each of two levels of illumination. These data show that increased contrast selectively displaces the time-velocity contour function in the expected direction: higher contrast enables the subject to perceive the edge of the moving stimulus at higher velocities with a constant value of t_1. Under low illumination the increase in contrast shifted the function from a to b, and under high illumination, the increase in contrast shifted the function from c to d.

Intensity level. The data plotted in Figure 7 also show an effect one would predict from statistical theory; namely, that contour perception of stimuli in motion is influenced by intensity. In Assumption 7 we stated that the mean slope of the decaying distribution of excitation became steeper as stimulus intensity increased. Furthermore, the rise in the mean slope of the distribution of excitation from the moving stimulus with increased intensity will be faster and will hasten the attainment of S_c. Accordingly, the importance of the summative effects of the initial stationary stimulus would be reduced.

As seen in Figure 7, the result of increasing intensity shifted the time-velocity contour function in the expected direction; under low contrast, increased intensity shifted the function from a to c, and under high contrast an increase in intensity shifted the function from b to d.

Maintaining a constant contrast between the stimulus and its surround while increasing intensity should and does facilitate contour perception, but only up to a point. If adjacent areas of a visual pattern differing in intensity have their intensities increased proportionately, a level will be reached where the necessary conditions for statistical peaking will no longer be possible. Thus if the areas defining a contour are both sufficiently intense, on-off receptors would not respond and the presumed normal distribution of excitation would degenerate.

Spatial displacement. Our account of how decaying neural activity from fixed stimulation combines with that produced by a moving stimulus requires particular spatial and temporal relationships of these two parts of the stimulation sequence. Any change which destroys the spatial and temporal contiguity of the two phases should have an adverse effect upon contour formation.

We predicted that the introduction of a spatial separation in the visual field between the position of the fixed stimulus and the initial position of the moving stimulus would require a higher value of t_1 to allow contour perception during movement than would be required in a control condition without spatial separation. The rationale for the separation in the visual field of the two phases of the stimulation sequence was to separate the assumed distributions of excitation produced by them in the visual cortex. It was necessary to provide the subject with a minute fixation point located at the place where movement began, because without it the subject probably would have tended to view the entire stimulus sequence with the same area of the retina. Following the cessation of the stationary stimulus, the moving stimulus appeared at the point of fixation and moved from left to right.

If the neural elements involved in the excitation produced by a stationary stimulus comprised a totally different population of elements from those involved in excitation from a moving stimulus, then no summation could take place. On the other hand, maximum summation would occur when identical elements comprised both populations. It was predicted that the greater the spatial displacement (the less the summation), the more adverse would be the effect upon contour formation.

Data obtained confirmed the spatial displacement hypothesis. Figure 8 presents the time-velocity functions obtained under several experimental conditions. These data are from eight subjects, and the procedure employed in obtaining the data was the same as that indicated previously. (Differences in the origin of the time-velocity functions which exist from one experiment to another are assumed to be due to individual differences, and differences in illumination and contrast.) The control condition consisted of having the stationary stimulus appear at the same place in the visual field from which movement began. Displacements were measured in degrees of visual angle at 1 and 2 degrees right, and 1 and 2 degrees left. Displacements of equal extent to the right and left produced equivalent shifts in the functions and so data were combined across this variable.

Temporal displacement. If there were a time lapse between the cessation of the initial stationary stimulus and the onset of the moving one, then subsequent summation or neural activity should be reduced. This reduction in summation would increase the time required to obtain S_c and would minimize the influence of the stationary stimulus upon the perception of contour of the moving phase.

Judgments of contour relating to this issue were obtained for eight subjects under each of three conditions. In the control condition the two phases of the stimulus sequence were temporally contiguous, where in the other two conditions temporal delays of 40

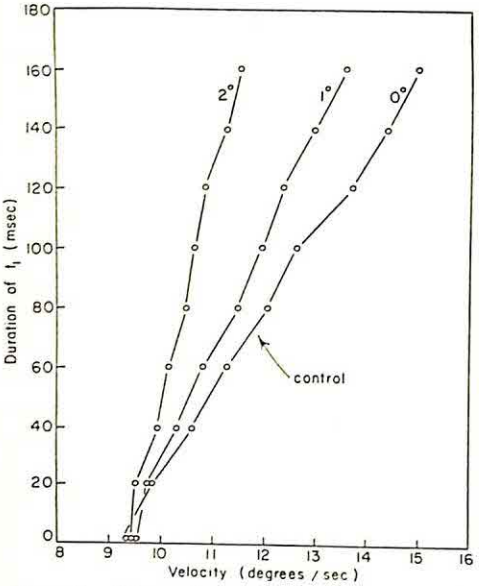


FIG. 8. The effects of spatial separation of the two phases of stimulation on dynamic contour. (Combined data of eight subjects. Stimulus brightness, 2.0 footlamberts; brightness of background, 70.0 footlamberts. Stimulus size, 0.5 degree. Extent of movement, 5 degrees of visual angle. Increasing the spatial separation between fixed phase of stimulation and moving phase required a greater exposure duration of the fixed phase in order to retain contour perception for a particular velocity of movement. Displacements to the right and left produced equivalent effects. Therefore, data were combined across this parameter.)

and 80 millseconds were employed. We predicted that the temporal delays would shift the time-velocity functions to the left indicating, that for a particular velocity, longer exposures of the stationary stimulus would be required to produce contour formation. The amount of neural activity available to sum with that produced by the moving stimulus is determined not only by temporal displacement of the two phases of the stimulus sequence, but by the duration of the stationary stimulus as well. Accordingly, we hoped to obtain data suggesting that what is lost to the summative process as a result of temporal delay can be recovered by increasing the value of t_1.

Data relevant to this problem are presented in Figure 9. The shift in the contour function as a result of temporal displacement is apparent. The greater the delay, the larger the shift to the left from the control condition. Presumably, a delay longer than 80 milliseconds might produce a time-velocity function asymptotic to the ordinate, suggesting that beyond a certain duration of temporal displacement, increased durations of t_1 would cease to influence the perception of contour. Increased values of t_1 *can* compensate for temporal displacement provided that the latter is not too long. For example, at a velocity of 18 degrees per second the duration of the sta-

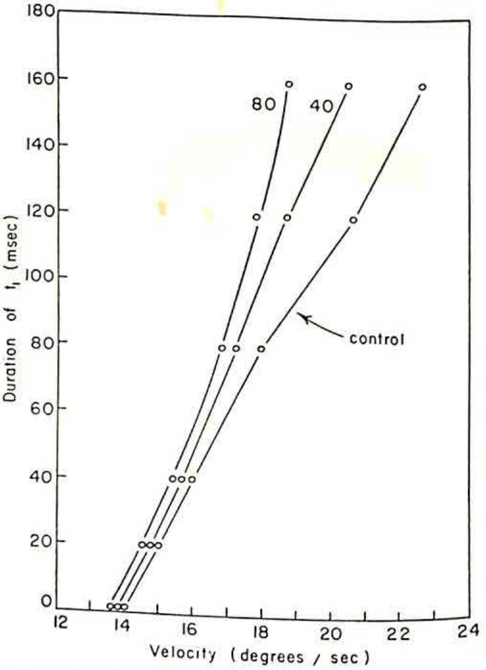


FIG. 9. The effects of temporal separation of the two phases of stimulation on dynamic contour. (Combined data of eight subjects. Brightness of stimulus, 2.0 footlamberts; brightness of background, 70.0 footlamberts. Stimulus size, 0.5 degree of visual angle. Extent of movement, 5 degrees. For a given velocity of movement, the exposure duration of the fixed phase of stimulation must be progressively increased as temporal separation increases in order to retain contour.)

tionary stimulus necessary for contour perception during movement was 80 milliseconds for the control condition (no temporal displacement). However, with temporal displacements of 40 to 80 milliseconds, the durations of the fixed phase had to be increased to approximately 100 and 120 milliseconds, respectively.

DISCUSSION

The conception of contour perception which we have offered, while congruent with certain facts concerning the visual system and with particular experimental results, has not directed itself to various considerations of a complicating and relevant nature. These considerations include the question of eye movement, the locus of the summation process, the question of neural inhibition, and some general questions about neurophysiological mechanisms.

Eye movements. The perception of edges of moving stimuli obviously raises the problem of how ocular tracking is related to the formation of contours. Surprisingly little empirical knowledge is available regarding the nature of ocular tracking, although one of the authors is currently investigating this problem. Having available a photoelectric system of measuring eye movements which is coupled to a system for generating moving targets, precise measures can be made of ocular tracking and these then related to perceptual judgments of targets whose velocity, displacement, and other attributes can be varied (Smith & Warter, 1959, 1960). Investigations to date indicate quite clearly that for the border of a moving target to be seen clearly a high degree of accuracy in visual tracking is *not* required. In fact, it appears that it is only for the very slow velocities (constant) of movement (about 10°/sec or less) that the eye tracks with reasonable accuracy. The upper limit of 10 degrees per second, it is important to note, is approximately the velocity of movement at which the borders of the moving stimulus begin to be seen as blurred when the duration of the stimulus in a fixed position before movement (t_1) is zero. It will be recalled that the most important single fact reported in our contour experiments is that by increasing the interval of exposure of the stimulus in a fixed position prior to movement we can systematically increase the upper limit of velocity at which the edge of the stimulus still can be seen clearly. A basic question raised by our experiments concerns how an increase in t_1 affects visual tracking. A provisional answer to this question based upon eye movement data to date is that increasing the magnitude of t_1 improves visual tracking generally but not to a level of accuracy characteristic of tracking at low velocities. About all we can conclude at this point is that the facilitation of dynamic contour perception by exposure of the target in a fixed position prior to movement is in some manner directly related to the accuracy of visual tracking. We can be sure that in our contour experiments the displacement of the stimulus at a constant velocity did not produce on the retina a displaced locus of maximum excitation which also was displaced at a constant velocity. We must assume that the cortical representation of a contour must contain components of acceleration and deceleration not found in the original stimulus. This, of course, is due to irregularities in the pursuit movement of the eyes under conditions wherein the moving stimulus alone is presented, and it is due to inaccurate tracking and the summation process under conditions in which both phases of the stimulation sequence are presented (see Figure 6).

Based upon eye movement data we can offer a suggestion concerning the possible relation between dynamic contour perception and correlated visual tracking movements. If one examines the magnitudes of lag and lead errors of the eye as the velocity of a stimulus increases, it is obvious that the magnitude of these errors increases as velocity increases, and at any given velocity it decreases at t_1 is increased. It is our hypothesis that as long as the lag or lead error does not exceed a few degree of angle (3–4 degrees at most), contour perception can be maintained. If this hypothesis is supported by the experiments now in progress, it will converge nicely with the facts about the maximum angular separation over which optional apparent movement can be obtained (De Silva, 1929), and the maximum distance over which one contour can exert an inhibitory influence over another (Fry & Bartley, 1935). Moreover, it is very interesting to note, and perhaps very important that the figure of 3–4 degrees angular magnitude is just about the visual field size represented by the measured size of receptive fields as revealed by Kuffler (1953).

The data obtained so far on the problem of eye movements tend to complicate the model, but do not vitiate it. Since the data on eye movements are at present incomplete we have assumed in the description of the summation process (see Figure 5) that the eye lags behind the target in the hypothetical situation depicted. If our eye movement data provide us with data contrary to this assumption, then the effect would be to augment the summation process by keeping the growth of neural excitation produced by the moving stimulus nearer the population of neural elements undergoing decay following the cessation of the stationary stimulus. Again, the important point is that the general statement of the model can accommodate the complicating factor of eye movements.

Locus of summation. Since the first statement of statistical theory by Marshall and Talbot (1942), they and subsequent statistical theorists have placed the locus of contour resolution in the striate cortex. The primary role of the projection system has been seen as contributing to the peaking of the distribution of neural excitation. The locus of summation of different distributions of excitation, however, has been the visual cortex alone. The question may be raised, however, as to the possibility of summation at lower levels. In an effort to ascertain the likelihood of this possibility we executed an experiment in which the stationary stimulus was presented to one eye whereas the moving stimulus was presented to the contralateral eye (Gulick, 1960). The results of this experiment suggest that for maximum contour facilitation it is necessary to present the stimulation sequence in such a way as to allow neural interaction and/or summation at the retinal level as well as at the cortical level. The implications of these data for the model involve a possible extension of the present view of neural summation as it is conceived in current statistical theory.

The interaction (and subsequent neural summation) of two temporally and spatially separate visual events which is central to our position raises in addition certain questions which require further investigation. We have presented evidence concerning the effects of separating in space and time the initial exposure of the stimulus in a fixed position and the movement component. It is necessary to ask what effects upon the perception of edge of the moving stimulus would be observed if the fixed and moving components of presentation differed in terms of certain attributes which have yet to be studied. For example, suppose that the stationary stimulus prior to movement were exposed out of focus, to be followed by the moving phase with the stimulus in focus. What then would be the change in our typical relation between velocity and t_1 as far as perceiving the moving edge is concerned? Similar important questions also need to be studied once we have completed the instrumentation necessary to present stimuli in this manner.

Neural inhibition and facilitation. As a final point of discussion we realize, as must the reader, that our conception of contour and its expression in neurological and physiological terms is unbalanced. The theory rests principally upon the concept of neurol summation and says little, if anything, about related concepts of inhibition and facilitation. In view of the inhibitory effects that one contour can have upon an adjacent one when both are fixed in position (Fry & Bartley, 1935), we are acutely aware that similar effects may be present when a particular contour is displaced relative to its original position. We think it reasonable, for example, to assume that an inhibitory mechanism can account for the "damping" of residual retinal activity correlated with the "trailing" edge of any moving stimulus. In any event we assume that future and more sophisticated expressions of statistical theory will incorporate modes of neural interaction in addition to summation at the cortex.

Regarding possible future expressions of statistical theory of contour we can indicate briefly some of the things we believe ought to be done. Additional perceptual studies are needed in which systematic investigations of the effects of differences in the fixed and moving components of the stimulus presentation are made. Moreover, we need to know more about the relationships between tracking motions of the eye and various conditions of stimulus movement. Finally, and most important, direct evidence bearing upon the neurological basis of perception is required if we are to continue to use with confidence the statistico-neurological assumption upon which the theory rests. While congruent with some neurological evidence, statistical theory of contour perception is far from being firmly based upon unambiguous neurological and physiological data. And, of course, the theory cannot find its real tests in perceptual data alone. We must continue to search for direct measures of the integrated activity of the visual system itself in relation to the data of perception. By so proceeding we will determine if statistical theory will prove itself as a fruitful means of conceptualizing basic problems in perception.

REFERENCES

BARLOW, H. B. Action potentials from the frog's retina. *J. Physiol.*, 1953, **119**, 58–68. (a)

BARLOW, H. B. Summation and inhibition in the frog's retina. *J. Physiol.*, 1953, **119**, 69–88. (b)

BARTLEY, S. H. *Vision: A study of its basis.* New York: Van Nostrand, 1941.

DAY, R. H. Application of the statistical theory to form perception. *Psychol. Rev.*, 1956, **63**, 139–148.

DAY, R. H. The physiological basis of form perception in the peripheral retina. *Psychol. Rev.*, 1957, **64**, 38–48.

DE SILVA, H. R. An analysis of the visual perception of movement. *Brit. J. Psychol.*, 1929, **19**, 268–305.

DEUTSCH, J. A. The statistical theory of figural after-effects and acuity. *Quart. J. exp. Psychol.*, 1956, **47**, 208–215.

FRY, G. A. *Blur of the retinal image.* Columbus, Ohio: Ohio State Press, 1955.

FRY, G. A. & BARTLEY, S. H. The effect of one border in the visual field upon the threshold of another. *Amer. J. Physiol.*, 1935, **112**, 414–421.

GRANIT, R., & SKOGLUND, C. R. Accommodation and autorhythmic mechanisms in single sensory fibers. *J. Neurophysiol.*, 1943, **6**, 337–348.

GULICK, W. L. Monocular contour perception under the influence of prior stimulation of the contralateral eye. *Psychol. Rec.*, 1960, **10**, 123–130.

HARTLINE, H. K. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Amer. J. Physiol.*, 1938, **121**, 400–415.

HARTLINE, H. K. The nerve messages in the fibers of the visual pathway. *J. Opt. Soc. Amer.*, 1940, **30**, 239–247.

KÖHLER, W., & WALLACE, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

KUFFLER, S. W. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.*, 1953, **16**, 37–68.

MARSHALL, W. H., & TALBOT, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. In H. Klüver (Ed.), Visual mechanisms. *Biol. Symp.*, 1942, **7**, 117–164.

MONNIER, A. M. *L'excitation électrique des tissues: Essai d'interpretation physique.* Paris: Hermann, 1934.

NERNST, W. Zur Theorie des elektrischen Reizes. *Pflüg. Arch. ges. Physiol.*, 1908, **122**, 275.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., & HEYER, A. W., JR. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

RUSHTON, W. A. H. The time factor in electrical excitation. *Biol. Rev.*, 1935, **10**, 1.

SKOGLUND, C. R. The response to linearly increasing currents in mammalian motor and sensory nerves. *Acta physiol. Scand.*, 1942, **4**, Suppl. No. 12.

SMITH, K. The statistical theory of the figural after-effect. *Psychol. Rev.*, 1952, **59**, 401–402.

SMITH, W. M., & GULICK, W. L. Visual contour and movement perception. *Science*, 1956, **124**, 316–317.

SMITH, W. M., & GULICK, W. L. Dynamic contour perception. *J. exp. Psychol.*, 1957, **53**, 145–152.

SMITH, W. M., & WARTER, P. J., JR. Photoelectric technique for measuring eye-movements. *Science*, 1959, **130**, 1248–1249.

SMITH, W. M., & WARTER, P. J., JR. Eye movement and stimulus movement: New photoelectric electromechanical system for recording and measuring tracking motions of the eye. *J. Opt. Soc. Amer.*, 1960, **50**, 245–250.

(Received August 23, 1960)

Psychological Review
1962, Vol. 69, No. 2, 109–135

SUBJECTIVE PROBABILITIES INFERRED FROM DECISIONS[1]

WARD EDWARDS

Engineering Psychology Laboratory, University of Michigan

Theories about how people make decisions in risky or uncertain situations have come to focus on two concepts: utility, or subjective value, and subjective probability. The status of the utility concept is fairly clear (for reviews, see Edwards, 1954c, 1961). But the status of the subjective probability concept is quite confused, both theoretically and in relation to experiments. Too often the concept of subjective probability has been introduced casually into a decision theory which focuses on the concept of utility. In fact, the only thorough treatment of a concept of subjective probability (quite different from the one to be discussed in this paper) is L. J. Savage's (1954) brilliant book. The notion of subjective probability would be necessary in psychology even if it played no role in decision theories, since men often make judgments about probabilities based on displays of random or nonrandom events. But this paper is mostly concerned with subjective probabilities inferred from decisions via some kind of value or utility maximization model of the decision-making process.

Psychologists trained in psychophysics tend to think that subjective probability is related to objective probability in more or less the same way that the subjective loudness of a tone is related to its objective intensity. The purpose of this paper is to examine the content, merits, and limitations of such an approach. The discussion will focus on two closely related matters. The first is the idea of a set of functions relating subjective to objective probability. The second is whether or not the subjective probabilities of a set of mutually exclusive events, one of which must happen, should add up to one.

The paper begins by defining two classes of decision theories. After some preliminary discussion of utility and subjective probability functions, it next considers the class of theories which result when subjective probabilities are assumed to add up to one. This class turns out to have some serious difficulties. A brief review of experimental evidence provides empirical reasons for avoiding these difficulties by rejecting additivity. Then the paper examines models which do not require subjective probabilities to add up to one. Such models require a reformulation of the concept of utility; utility scales must have true zero points. Finally the paper derives two experimentally testable properties which any plausible utility-subjective probability model, additive or otherwise, must have, and presents a model which I find attractive.

---

[1] The work reported here was supported in part by the Operator Laboratory, Air Force Personnel and Training Research Center, and in part by Project MICHIGAN under Department of the Army Contract DA–36–039–SC–78801, administered by the U. S. Army Signal Corps. Reproduction for any purpose of the U. S. Government is permitted. Credit amounting to coauthorship should go to L. J. Savage, W. M. Kincaid, and George Minty, who did the mathematical work on which this paper is based; their individual contributions are acknowledged where they occur. I am also deeply grateful for advice and criticism to them and to many others, especially G. M. Becker, C. H. Coombs, R. D. Luce, Jacob Marschak, Anatol Rapoport, M. Toda, E. L. Walker, and K. V. Wilson. An earlier version of this paper (Edwards, 1959, 1960) is closely related.

MAXIMIZATION AND UTILITY

First, some definitions and assumptions. This paper is concerned with decisions of the following kind. A decision maker can choose one and only one of two or more courses of action. Associated with each course of action is a finite set of possible outcomes, one and only one of which will happen if that course of action is chosen. Each possible outcome can be described by a number, called a utility, defined at least up to a linear transformation (i.e., measured at least on an interval scale); this number fully represents the desirability of that outcome to the decision maker. If more than one possible outcome is associated with a given course of action, then a well defined external event, over which the decision maker has no control, determines which outcome will occur if that course of action is chosen. Most nonsocial decisions fit this description.

The crucial assumption of the class of theories considered here is that people behave as if they were attempting to maximize some quantity—perhaps not skillfully, perhaps not consistently, but still attempting to maximize. Maximize what? The following equation defines the concept of subjectively expected utility (SEU):

$$\mathrm{SEU} = \sum_i \psi_i u_i \qquad [1]$$

It refers to a course of action which has a number of possible outcomes. The i^{th} possible outcome has an objective value to which corresponds a subjective value or utility u_i. Outcome i will occur if a given external event occurs; that event may have an objective probability, or the objective probabilities of one or more of the relevant set of external events may be undefined. In either case, outcome i has a subjective probability ψ_i. The theories with which this paper is concerned simply assert that subjects (*S*s) behave as though they choose, from among the courses of action open to them, the one which has the largest SEU. For more extended discussion of the origin of these models, see Edwards (1954c, 1955). This paper will treat them as models of what men do (descriptive models) rather than of what they should do (normative models).

Equation 1 introduces formally the central concept of this paper: subjective probability. What is a subjective probability? It is a number between zero and one which describes a person's assessment of the likeliness of an event. Exploration of several more explicit definitions follows.

This paper assumes that objective probabilities exist, are known for some events, and are distinct from subjective probabilities, though subjective probabilities are related to objective ones. Some theorists like Savage (1954) have asserted that there is no such thing as objective probability; all probabilities are subjective. This is no place to argue whether or not objective probabilities can be meaningfully defined. The argument is irrelevant. Some events, like the toss of a coin or the roll of a die, have well established "conventional" probabilities. Whether these probabilities are objective or not, they can meaningfully be compared with the subjective probabilities revealed when people make decisions which involve these events.

Of course many events exist which

are neither certain nor impossible but for which no objective or conventional probability can be defined. Example: Within the next 24 hours, you will drink a glass of wine. Such events are called *uncertain*, while events for which objective probabilities can be defined are called *risky*. This paper will not consider uncertain events. The same decision theory ought to apply to decisions whose outcomes are contingent on uncertain events as to decisions whose outcomes are contingent on risky events, and the latter are more helpful in examining the nature of that theory.

The notion of a set of functions relating subjective to objective probability is vital to this paper. But the theorems proved below are relevant also to theories which deny or ignore the existence of such a relationship. They imply for such theories the existence of severe restrictions on permissible relationships between subjective probabilities and classes of risky events (classes defined by objective probability values).

From here on, some abbreviations are needed. The word "maximization" will be omitted from phrases like "the SEU maximization model." The concept of objective probability will be symbolized OP in text; SP will stand for subjective probability. The concept of a set of events of which some one must happen and not more than one can happen—that is, the concept of an Exclusive, Exhaustive set of Events—will be symbolized EEE in text. Two classes of SEU models exist. Those in which the SPs of an EEE must add up to a constant will from here on be called ASEU (additive subjectively expected utility maximization) models. Those in which the SPs of an EEE do not need to add up to anything in particular (though of course the OPs of an EEE must always add up to one) will be called NASEU (nonadditive subjectively expected utility maximization) models.

The subject of transformations will come up. Psychologists speak of three classes of scales, defined by three transformation groups, which are relevant here. An interval scale permits the linear transformation $x' = ax + b,\ a > 0$. A ratio scale permits the scale transformation $x' = ax,\ a > 0$. An identity scale permits only the identity transformation $x' = x$. This paper will use the psychological rather than the mathematical names for these classes of scales.

CLASSES OF SEU MODELS

This section will identify two main classes of SEU models, ASEU and NASEU, and will then examine some extraordinary restrictions on the utility functions which must enter into each. The difference between ASEU and NASEU models is defined by

Theorem 1: If only interval, ratio, and identity scales are considered, in any SEU model in which the SPs of an EEE must add to a fixed constant (ASEU model) utility must be measurable at least on an interval scale and SP must be measurable at least on a ratio scale. In any SEU model in which no restrictions are put on the sum of the SPs of an EEE (NASEU model), utility and SP must both be measurable at least on ratio scales.

Consider the NASEU models first. Since the transformations under consideration are successive restrictions of the general linear transformation, consider the least restrictive possibility first. In that case, the transformation $u' = au + b,\ \ a > 0$ on utility and $\psi' = c\psi + d,\ c > 0$ on SP are permissible. Consider two courses of action, A and B, which have equal SEUs. (Continuity assumptions not spelled out here are necessary to guarantee that such events exist and that they do not have certain pathological properties. Axioms 1 and 2 of the mathematical Appendix, plus the assumption that there are at least three possible outcomes with different utilities, would be more than enough.) Permissible transformations must not change equal SEUs, which imply indifference between A and B, to unequal SEUs, which imply that one is preferred to the other. Consequently:

$$\sum_i \psi(A_i)u(A_i)=\sum_j \psi(B_j)u(B_j)\leftrightarrow$$
$$\sum_i \psi'(A_i)u'(A_i)=\sum_j \psi'(B_j)u'(B_j) \quad [2]$$

A_i = i^{th} possible outcome of course of action A, $i = 1, 2, \cdots, I$.

B_j = j^{th} possible outcome of course of action B, $j = 1, 2, \cdots, J$.

Substituting the transformations under consideration into the second half of Equation 2 produces:

$$\sum_i [(c\psi[A_i]+d)(au[A_i]+b)] = \sum_j [(c\psi[B_j]+d)(au[B_j]+b)]$$

Multiplying and collecting terms:

$$ac[\sum_i \psi(A_i)u(A_i) - \sum_j \psi(B_j)u(B_j)] + bc[\sum_i \psi(A_i) - \sum_j \psi(B_j)] + ad[\sum_i u(A_i) - \sum_j u(B_j)] + bd[I - J] = 0 \quad [3]$$

By hypothesis:

$$\sum_i \psi(A_i)u(A_i) - \sum_j \psi(B_j)u(B_j) = 0$$

and so the first term of Equation 3 vanishes. Since both b and d may assume any arbitrarily chosen values, the only way to guarantee that the sum of the other three terms will be zero for any values of b and d is to require that the quantity inside the brackets in each term be zero. But the quantity inside the brackets in the second term (and in the third also) is explicitly not required to be zero. Consequently this set of transformations cannot be permissible. Since both a and c are required to be greater than zero, the only way to guarantee that Equation 3 will be correct is to require that both b and d be zero. That, of course, means that both utility and SP must be measured on ratio scales. Thus the second half of Theorem 1 is established.

Now consider the first half of Theorem 1, which is concerned with ASEU models. In such models, the first and second terms of Equation 3 are zero by hypothesis. Furthermore, it is clear that in such models SPs may not have an arbitrary origin, since they are required to add up to a fixed constant, and consequently d must be zero. If d is zero, the other two terms vanish, and no restrictions on b are necessary. This establishes the first half of Theorem 1.

Possible Relations between Utilities and Objective Values

Luce (1959) has recently pointed out some surprising and severe restrictions on possible relationships between objective and subjective continua of any sort, in a paper of extraordinary depth and importance. He asserts two principles to which any function relating an independent to a dependent variable within a theory should conform. First:

> Admissible transformations of one or more of the independent variables shall lead, via the substantive theory, only to admissible transformations of the dependent variables. (p. 85.) (Note that a form of this principle was used to prove Theorem 1.)

Secondly:

> except for the numerical values of parameters that reflect the effect on the dependent variables of admissible transformations of the independent variables, the mathematical structure of the substantive theory shall be independent of admissible transformations of the independent variables. (p. 85.)

The first condition means, for example, that a theory in which the dependent variable is measured on a ratio scale may *not* require that when the arbitrary origin of an interval scale independent variable is changed, the origin for the dependent variable must change also. The second, more subtle, condition means that the mathematical form of the function relating independent to dependent variables may not be changed from, for example, a power function to a simple sum of variables by permissible transformations on the independent variables. Both of these principles of theory construction seem plausible and even necessary. What do they imply for utility and SP?

First, consider interval scale utility as dependent variable and ratio scale money (or quantity of any other continuously variable good) as independent variable. The relation between these two variables, if it is to conform to the two requirements of the previous paragraph, must, according to Luce's theorems, have the form $u(\$_i) = a \log \$_i + b$, where a and b are arbitrary constants and $\$_i$ is any amount of money. If, as is necessary in NASEU models, utility is measured on a ratio scale, the relation between it and money must, according to Luce's theorems, be $u(\$_i) = a\$_i^b$ (again, a and b are arbitrary constants).

These restrictions on the possible forms of utility functions are astonishingly severe. For example, they imply that if for interval scale utilities there are any points on the function relating utility of money to its objective value at which that function increases (that is, if a is positive), and if the value of b is finite, which it must be if the equation is to make sense, then the utility corresponding to zero dollars (or some fixed positive or negative amount, since the origin may be displaced by a fixed amount) is minus infinity. The utility of negative amounts of money (or amounts less than the fixed amount) is undefined. And the form of the utility function is necessarily negatively accelerated; that is, it is impossible to find a person who attaches as much as twice as much utility to $\$2x$ as he does to $\$x$. In short, interval scale utilities associated with ratio scale money values require ridiculous utility functions.

The relation between ratio scale utilities and ratio scale money values is much less ridiculous. Its most important restrictive characteristic is that it may have only one inflection point. That means that if the function is ever positively accelerated, it is always positively accelerated; if it is ever negatively accelerated, it is always negatively accelerated. Even this characteristic is pretty restrictive. But it is not utterly ridiculous, as are the properties of interval scale utilities discussed in the previous paragraph.

These restrictions are so surprising that one is naturally tempted to try to wiggle out of them by arguing against the assumptions which lead to them. Luce (1959) has extensively discussed these assumptions and their possible limitations. This paper will not cover the same ground again; instead, I will simply record my opinion that at least for interval scales Luce's restrictions are inevitable and we must therefore learn to live with them.

Do these principles of theory construction place any restrictions on possible relations between SP and OP? Probably not. We usually consider OP to be measured on an identity scale; if it is, Luce's arguments place no special restrictions on the mathematical form of possible relationships between it and anything else.

GENERAL CHARACTERISTICS OF PSYCHOPHYSICAL DECISION MODELS

Before getting down to specific issues concernings SPs, I must set the stage by discussing some general issues and some characteristics which all ASEU or NASEU models must or should have.

Identifying Characteristics

To predict choices from an SEU maximization model of any kind, you must know the utility and the SP of each possible outcome of each available course of action. You can then substitute these values into Equation 1, perform some arithmetic, and predict that the course of action with the highest SEU will be chosen. Such predictions will be wrong on occasion. This fact can be included in the model (at a heavy penalty in additional mathematical complexity) by making it stochastic, or it can be excluded from the model and left to a general theory of errors, as has been customary in most of psychophysics. This paper does the latter, but it seems likely that many of the conclusions established here for deterministic models will apply to stochastic models of the same type as well.

In order to use an SEU model, then, you need to know utilities and SPs independently of the decision you are trying to predict. How can you find them? That is a very difficult, much debated question. For the purposes of this paper, it is enough to assume that some psychophysical method is available which gives accurate utility and SP measures for any object or event of interest, and that you have available such information about anyone whose decisions you are predicting.

In what form can such information be available? Consider first information about utilities. If money is the valuable object in question, presumably a table could be prepared showing the amount of utility an *S* attaches to a gain of one dollar, of two dollars, of three dollars, and so on. (My reasons for saying "a gain of one dollar" rather than simply the amount one dollar or a total fortune of $5001 will become clearer later; the differences among these formulations are irrelevant at this point.) But the information in that table would have to be based on experiments in which the *S* received or lost particular dollar bills, with particular serial numbers. How do we know that he will respond similarly to different dollar bills, or perhaps even to nickels and dimes instead? Of course we don't. We simply assume that (within reasonable limits) the form in which *S* receives the increment of one dollar is irrelevant; that is, we abstract the property "United States currency with an objective value of one dollar" from a variety of other properties of the stimuli used in the utility measurement experiment, and assume that it is the one on which the results most significantly depend.

In the case of utility, the point made above is so familiar as to be trivial. But what about the case of SP? On what properties of events do SPs most significantly depend? The most obvious characteristic of events which should turn out to be highly correlated with SP is, of course, OP (remember that this paper is restricted to classes of events for which OPs are well and independently defined). Unfortunately, unpublished data make it perfectly clear that if two bets are identical in payoff, OP, and all other characteristics except the means by which the OP is displayed, *S* may markedly prefer one to the other—so much so that he is willing to pay for the privilege of using his preferred display, even though he loses in expected value (hereafter abbreviated EV) by doing so. Within the framework of the theories discussed in this paper, this finding can be explained only by assuming that several different SPs may correspond to a given OP, depending on the nature of the OP display. In short, any attempt to state SP as a function of OP alone is doomed to failure.

What then is SP a function of? Mathematical decision theorists, aware of logical difficulties which result from thinking of SP as a function of OP, often use a quite different approach. They think of SP as a characteristic of interaction between a specific event and a specific person. This, of course, means that in order to make a prediction they must independently determine the SP associated with the particular event about which the prediction is made for the person whose behavior is to be predicted. Such a strategy for theorizing has two major difficulties. The first is that it makes little attempt at generality. Obviously some characteristics of an event are important in determining its SP and some are not; we need to know which are which. Furthermore, if SPs are to be used to predict real-world decision making behavior (e.g., for designing displays in military information processing systems), some set of rules which govern the relationship between event and resulting SP must be found so that predictions can be made without running experiments on each situation and person to which the predictions are intended to apply.

What we need and do not have is some systematic empirical information about various displays of probabilities. In the absence of identified stimulus dimensions relevant to SP other than OP, decision theorists are in as awkward a position as psychologists interested in loudness would be in if nothing were known about the physics of sound. Of course this position can only be made worse by ignoring the one stimulus variable which we know to be highly correlated with SP and therefore important: OP. But OP is not enough. Decision theorists cannot much longer get along without empirical information about the psychophysics of probability displays.

Since we have no useful information about displays of probabilities, this paper can only assume that they are important but leave their effects unspecified. The following discussion assumes that each event to which a decision theory might be applied has certain as yet unspecified characteristics which determine its SP. One of those characteristics is, of course, OP, whenever OP is defined. The others are unknown; this paper will call them identifying characteristics. Only two assumptions about identifying characteristics will be consistently made. The first is that whenever an event is repeated as exactly as possible, its identifying characteristics are unchanged; in other words, identifying characteristics are recurrent, not unique. (The contrary possibility would make them inaccessible to scientific study.) The second assumption is that it is quite possible to have events with different identifying characteristics but the same OPs; this simply means that OP is not perfectly correlated with identifying characteristics. (If it were, then Theorem 3 below would imply that SP always equals OP, and so the point of the notion of identifying characteristics would be lost.)

The SP Function Book

Mathematically, this paper will conceive of SP as a function of OP and of identifying characteristics. One way to think about this function is to imagine a book. Each page of this book has on it a page number (the quantity *r* of the Appendix) and a function which relates SP to OP. These functions need not necessarily be continuous. The page number is simply a code which refers to a particular set of identifying characteristics. Thus, given any event, you can find its SP by finding out its identifying characteristics, looking up the number corresponding to that particular set of identifying characteristics, finding the page in the book which has that page number, and then looking up the SP corresponding to the OP of the event in question on that page. Mathematicians may find this concept and the notation of identifying characteristics clearer if they examine the axiom system and proofs of Theorems 2 and 3 in the Appendix. The assumptions of that proof formally define ASEU models of this class. It turns out that the formulation in terms of a function book, even if that book may have an infinite number of pages, implies that SP always equals OP if additivity of SPs is assumed. But the notion of an SP function book is appropriate for NASEU models, for which no such implication exists.

The idea of the previous paragraph should be familiar to psychophysicists. This model is analogous to a familiar model in auditory psychophysics. Perceived loudness of pure tones is a function of both stimulus intensity and stimulus frequency. A number of different functions relating loudness to intensity, each using a different value of frequency as a parameter, are all an auditory psychophysicist would need to make predictions about loudness. (He would actually use equal loudness contours for the purpose, but the underlying model is the same.)

The purpose of this notion is to provide an orderly way to deal with the possibility that more than one SP may correspond to a given OP, depending on display characteristics. Of course, I hope that such a book of SP functions would have relatively few pages, and an orderly structure, so that the sequence of pages is not arbitrary. But neither of these characteristics need be assumed in the following discussion.

Incidentally, uncertain events as well as risky ones can be provided for within the notion of a book of SP functions. Presumably uncertain events will have identifying characteristics different from risky events (indeed having no definable OP is presumably itself one identifying characteristic), and so pages appropriate for such events can be segregated in one section of the book. Each such page has, as its function relating SP to OP, a flat horizontal line running from $OP = 0$ to $OP = 1$. (More precisely, this line is defined only for $0 < OP < 1$.) Thus once its identifying characteristics are known, the SP of an uncertain event is completely specified.

ASEU MODELS

In spite of major differences in the axiom systems which lead to them and in the philosophy which motivates them, for the purposes of this paper most of the ASEU models are very similar mathematically and operationally to expected utility maximization models (hereafter called EU models) like the one proposed by von Neumann and Morgenstern (1947). The only difference, and it is a crucial one, is the use of SPs instead of OPs in the models.

Finite vs. Infinite Event Sets

Two main classes of possible ASEU models are defined by the dichotomy between finite and infinite event sets. Mathematically, the distinction is simple enough. If we consider all possible outcomes of 10 tosses of a coin, then this is a finite event set. If we throw a dart at a target, then all possible places it can hit form an infinite event set (note that the point of the dart need *not* be a mathematical point with zero area). Obviously a person can encounter only a finite set of events in his lifetime, but that is irrelevant. The point is that he will encounter many situations in which some one of an infinite set of possible events may occur. At first glance, it would seem desirable, then, to have a decision model capable of coping with any member of an infinite event set—and most decision models are just that. A few theorists (notably Davidson, Suppes, & Siegel, 1957) have preferred to think about finite event sets, mostly because they wanted to be able to check all of their predictions, instead of some finite subset of an infinite set of predictions. As will become clear later, finite event sets immensely complicate the mathematics, and at the same time reduce the value of the model by making it inapplicable to situations in which the set of possible events is infinite. Although this paper discusses finite models below, I consider such models far less interesting than the infinite models.

Please remember that the notion of an infinite event set is not the same thing as the notion of a continuum of OPs. If all elements of a continuum of OPs are realizable by members of an event set, then that set is infinite (nondenumerably infinite, to be precise), but an infinite set of events, all with different OPs, may still have the property that between any two OPs which are associated with members of the event set lies an OP which is not. In fact, continuity of OPs is not assumed directly in Theorems 2 and 3 below, though it can be deduced as a theorem from the nonatomicity property which is assumed.

Restricted vs. Unrestricted Sets of SP Functions

How many different SPs may be associated with different events all having the same OP? Or, to ask essentially the same question in a different form, how many different functions may relate SP to OP? Still another formulation: how many pages may there be in the SP function book? Three classes of answers to this question exist for infinite event sets, and two for finite ones. For infinite event sets, the number of different SPs which may be associated with different events all having the same OP may be finite or infinite. If that number is infinite, it may be denumerable (capable of being set in one-to-one correspondence with the integers) or nondenumerable (the set of all points on a line segment is nondenumerable; so is the set of all OPs). It turns out that the finite and denumerably infinite cases have the same properties, so only the latter, less restrictive case will be considered.

The formulation of the relation between OP and SP in terms of the SP function book tacitly presupposes that the set of different SPs which may be associated with different events all having the same OP is at most denumerably infinite. If, as the preceding discussion has suggested, the difference between two different events with the same OP but different SPs lies in the way in which that OP is displayed, then the assumption of denumerability is simply the assumption that all possible ways in which an OP could be displayed could be listed (in infinite time). It is difficult to see why anyone would resist so mild an assumption—and yet Theorem 3 will show that its consequences are drastic. Nevertheless, it seems greedy for a theorist to insist that a denumerably infinite set of different SPs to associate with a given OP is not enough and that he wants more. I prefer to accept denumerability and abide by its rather drastic consequences.

For mathematical convenience, however, it is expedient first to examine a restriction wich applies even to the nondenumerable case. It is well known that an infinite number of different probability measures can be defined on a given event set. This makes it seem unlikely that any substantial formal restrictions exist on models of the nondenumerable "anything can happen" sort. But there is at least one. Usually SPs rise with OPs. In fact, it is tempting to assume that they always do so. Consider two events, A and B. If the OP of A is higher than the OP of B and the SP of A is less than the SP of B, call that pair of events an *inversion*. The assumption that SPs always rise with OPs is an assumption that inversions cannot occur. But unless inversions exist in infinite numbers and at all values of OP, then it is easy to prove that SP always equals OP. Theorem 2 below states a stronger assertion from which this one is a self-evident consequence.

Theorem 2: Assume that an infinite, nonatomic event set is being considered, that every event has an SP and an OP, that the SPs of any EEE must add to one, and that some event can be found for which SP $\neq$ OP. It follows that for any OP p, $0 < p < 1$, two events A and B can be found such that $p(A) = p$ and $p(B) < p(A)$ and $\psi(B) > \psi(A) + Q$ where Q is a noninfinitesimal quantity (unless the maximum difference between SP and OP is infinitesimal, or $p(A)$ or $p(B)$ is infinitesimally different from 0 or 1).[2]

A more precise mathematical formulation of Theorem 2, and its proof, are long, technical, and contribute little to the following discussion. I have therefore put them in a mathematical appendix which follows the list of references. The only important assumption necessary to prove Theorem 2 which has not already been discussed is that of nonatomicity; a discussion of it follows Theorem 3, for which it is also essential. Its content is primarily technical. The assumption that all events in the event space have both OPs and SPs has been discussed. The assumption that some event exists for which SP $\neq$ OP is the only reason for even thinking about SPs in risky situations; without it there is no point in a theory about SPs.

Lemma 4 of the mathematical appendix implies some further statements about subjective probabilities in this nondenumerable case. If SP is ever different from OP, there must be some event or events for which the size of that difference (disregarding sign) is a maximum, or at any rate is only infinitesimally different from a maximum. In fact, because of the additivity of SPs there must be at least two such events, one in which the difference is positive and one in which it is negative. Call the event with the smallest OP for which this maximal difference is present and positive M. Event M has an OP $p(M)$ and an SP $\psi(M)$. For any Event A, define $d(A) = \psi(A) - p(A)$. Then $d(M)$ is the largest positive value of $d(A)$ that can occur, and $d(\bar{M})$ is the largest negative value of $d(A)$ that can occur. Now what can be said about the difference between SP and OP for events other than M? For any such Event A, max $d(A) = d(M)$, and min $d(A) = d(\bar{M})$. Lemma 4 shows that for every probability p there must be an Event B having OP p such that:

---

[2] This theorem and its proof were formulated by W. M. Kincaid.

$$d(B) = [d(M)]\left[\frac{q(B)}{q(M)}\right]$$

if

$$p(B) \geqq p(M)$$

and

$$d(B) = [d(M)]\left[\frac{p(B)}{p(M)}\right]$$

if

$$p(B) \leqq p(M)$$

There must also be an Event C having OP p such that:

$$d(C) = [d(\bar{M})]\left[\frac{q(C)}{p(M)}\right]$$

if

$$p(C) \geqq q(M)$$

and

$$d(C) = [d(\bar{M})]\left[\frac{p(C)}{q(M)}\right]$$

if

$$p(C) \leqq q(M)$$

Finally, it shows that events must exist having that same OP p and all possible values of ψ between $\psi(B)$ and $\psi(C)$. In other words, in a space relating SP to OP there is a parallelogram of points defined by the four vertices $(0, 0)$ $(1, 1)$, $[p(M), \psi(M)]$, and $[p(\bar{M}), \psi(\bar{M})]$. All points within this parallelogram must be realizable by events in the event space with appropriate subjective probabilities.

The preceding paragraph leads up to Theorem 3. The awkward property exhibited in Theorem 2 applies to models which permit a nondenumerably infinite set of SPs to be associated with any given OP. Next consider what happens if any OP can be found (other than 0 or 1) with which no more than a denumerably infinite set of SPs is associated.

Theorem 3: As in Theorem 2, assume an infinite, nonatomic event set, each event having an SP and an OP, and assume that the SPs of any EEE must add to one. In addition, assume that there is some OP other than 0 or 1 which has the characteristic that the number of different SPs which occur in conjunction with different events all having that OP is at most denumerably infinite. It follows that SP always equals OP.[3]

Again a more precise formulation of the theorem, and its proof, are in the mathematical Appendix. Certain additional assumptions of purely technical importance (e.g., continuity at zero) are necessary to the proof, but all the assumptions which I consider at all controversial are stated here.

At this point it is worth while considering the assumption of nonatomicity which underlies Theorems 2 and 3. Suppose I assert "For any event which is already included in an event set, I wish to include another which is defined as the simultaneous occurrence of the event in question and heads in the toss of a coin which I happen to have handy." This assertion defines a nonatomic event set (in fact, it is somewhat stronger than is necessary to specify a nonatomic event set). In other words, the property of nonatomicity simply asserts that any given subset may be further subdivided. Unless the set of events with which you intend to work is rigidly defined in advance, it is impossible, as I see it, to think of any real world event set which is not nonatomic. However, this property is far from trivial. For example, the assumption of nonatomicity directly implies that every OP is realizable by some event in the event set, and that the event set is nondenumerably infinite.

---

[3] This theorem and its proof were formulated by L. J. Savage. W. M. Kincaid turned Savage's sketched proof into the form contained in the Appendix.

All in all, the characteristics of ASEU models based on infinite classes of events seem unattractive. We may adopt any of three strategies. One is to accept and live with these unfortunate properties. That possibility has already been examined. It is now time to mention very briefly the second set of possibilities.

Finite Event Sets

The arguments given previously against finite events sets were, I think, strong enough to argue against their serious consideration. However, it is interesting to speculate further about their mathematical properties. It seems quite likely to me that analogues of Theorems 2 and 3 must exist for finite cases. Such theorems would define maximum differences between SP and OP as a function of number of events in the finite event set, spacing of the OPs of those events on the OP continuum, number of different SP functions permitted (if that number is less than the number of events in the event set), and number and location of inversions of ordering between SPs and OPs. It seems plausible both that such theorems exist and that Theorems 2 and 3 are limiting cases as the number of events in the event set approaches infinity. Unfortunately, I have been unable even to guess what the form of such theorems might be (except for specially constructed cases, such as the case of equal spacing of events along the OP continuum). As is often the case, it looks as if the finite cases are mathematically much more difficult than the infinite cases. Even in the case where only a single monotonic function relating SP to OP is permitted (surely the simplest finite case), I have been unable to formulate a specific theorem relating number and spacing of events in the event set to maximum possible difference between OP and SP.

The Third Strategy

This paper has considered two classes of ASEU models based on infinite event sets and a class based on finite event sets; all are unattractive for various reasons. The third possibility is to abandon ASEU models altogether and use NASEU models instead. Some previously discussed considerations about utility functions make this option attractive. But the NASEU models are so much less powerful than the ASEU models that I would be very reluctant to turn to them on a priori grounds alone. So before examining them seriously, it is worth while to review the empirical evidence which bears on them.

THE PROBABILITY PREFERENCE DATA

A rather substantial body of data bears on the relation between SP and OP. These data come from a series of experiments (Edwards, 1953, 1954a, 1954b, 1954d, 1955) which I have collectively called the probability preference experiments.

The basic design of the probability

preference experiments was as follows: Lists of bets were prepared. Each list included eight bets, with OPs of winning (or losing, depending on the list) running from 1/8 through 8/8. In various experiments a total of eight different lists were used. The eight bets in each list all had the same objective EV, so that there was no objective reason to prefer one bet in a given list to any of the others in the same list. All bets in each list were paired with one another according to the method of paired comparisons. At various times, a total of well over 100 college students were required to choose between the members of these pairs. In some experiments, the *S*s sat in classrooms, looked at slides of the pairs of bets, and made choices "as if they were gambling." Other *S*s, run individually, first made imaginary choices, then gambled for worthless chips, and finally for real money.

The main results of these experiments, taken as a whole, were:

1. Although there were substantial individual differences in choices, certain patterns of choices showed up in all experimental groups and in just about all individuals. The two most outstanding of these patterns were that *S*s usually preferred the bet with the 4/8 probability of winning from any positive EV list to the other bets on that list, and usually preferred the bet with the lower probability and higher amount of loss in any pair of negative EV bets.

2. The preferences cited above, and indeed the complete pattern of choices, were relatively independent of EV level of the list of bets involved, so long as the zero point was not crossed. This means that the preferences observed should be attributed to the OPs involved, which were constant, rather than to the amounts of money, which varied from list to list. This fact suggests that for these students and these amounts of money, the utility curve is relatively linear with amounts of money, while the SP curve is not linear with OP. It also suggests that the value of OP from which SP deviates most widely is 0.5, for positive EV bets. Using the assumption that SP = OP and the further assumption that the size of the just noticeable difference for utility is half a cent, it is impossible to construct a utility function to account for the preferences observed. If the size of the just noticeable difference for utility is assumed to be zero, of course such a curve can be constructed, but it has at least 12 inflection points between \$0 and +\$6. The same sorts of statements can be made if the data are analyzed *S* by *S*.

3. The complete pattern of choices changed radically from positive EV to negative EV bets, even though exactly the same OPs were used. The main change was a strong preference for negative EV bets in which the probability of losing was low and the amount of possible loss high. When this preference was removed from the data by crude statistical means, the residual preference pattern was pretty much the mirror image of the preference pattern for positive EV bets. These findings suggest that there is a strong interaction between utility and SP.

4. The differences among just imagining, gambling for worthless chips, and gambling for real money are discernible, but not sufficiently serious to invalidate any of the conclusions discussed above. This means that the difficulties which might arise in a gambling experiment because *S* wins or loses on each bet before he decides about the next bet are much less severe than a priori considerations might lead one to expect.

5. Minor changes in wording, in the choice of payoff events and in the financial status of *S*s do not make any detectable differences in their choices. In particular, *S*s can be made to win exceedingly large amounts of money, or made to lose substantial but smaller amounts of money, without significantly changing their patterns of choices. This finding has been confirmed in a study devoted especially to it (Edwards, 1954d).

Findings 1 and 2 cause me to reject the hypothesis that SP = OP. Therefore, they, in conjunction with the considerations discussed in the previous section, cause me to reject all EU and ASEU models. The implications of Findings 3, 4, and 5 for a NASEU model are important; they will be discussed below.

NASEU MODELS

What are the mathematical properties, if any, of a NASEU model? Take a look at Equation 1. It defines NASEU as well as ASEU models. But what does the operation of addition called for by the sigma in Equation 1 mean for a NASEU model? For an ASEU model, that sigma simply derives from the usual definition of a probability measure, and reflects the additivity property of that measure. For a NASEU model, it has no such simple mathematical justification. A variety of psychological arguments about its plausibility are available. This paper will not press them; evidence to be presented later argues against this form of NASEU model, and I do not advocate or defend such a model. The purpose of this section is to introduce some ideas which are a part of the model which I do defend. For the time being, therefore, it is enough to say that no law of men or mathematics forbids the multiplication of utility and nonadditive SP values, the algebraic addition of the products, or test of the hypothesis that people make decisions in such a way that they in fact maximize the resulting sums.

Ratio Scale Utilities

However, Theorem 1 indicates that a NASEU model must use a ratio scale measure of utility. A ratio scale of utility implies that utility has a true zero point. Where is it? Only one answer is plausible: where you now are. Zero utility is your current position, and you can never leave it.

This is not exactly a new idea. The first implicit use of it was by Mosteller and Nogee (1951), who used the EU model to make a utility scale for money (and, incidentally, proposed nonadditive SPs as an alternative explanation for their results). Mosteller and Nogee used a gambling situation in which real money changed hands during the course of the experiment. But in determining the utility of a given amount of money, they simply found an indifference point which involved that amount of money without moving the origin to take into account *S*'s financial status at the time they measured it. Such a procedure can yield a classical utility function only if the form of that function is invariant up to a linear transformation under movements of the origin (through which all utility functions conventionally pass) along the function. Only a very limited class of functions, of which the most familiar member is the straight line, has this property, and the utility curves Mosteller and Nogee obtained don't look like any member of that class. Consequently those curves in fact defined zero utility as the present monetary status of each *S*, and so were of the type that this section advocates.

The first self-conscious use of an idea like that proposed here was made by Markowitz (1952). He suggested that the zero point of the utility scale be taken as *S*'s customary financial position, and that the form of the function changes when that customary financial position changes. He used this idea to remedy a deficiency in Friedman and Savage's (1948, 1952) previous account of gambling and insurance buying. The only difference between Markowitz's position and this one is that Markowitz defines zero utility as the customary position, while this paper defines it as the current position.

Although this conception of utility is novel and quite different from the traditional one, it is neither internally contradictory nor absurd. Nor does it imply that people cannot change their habits as they change their financial status. Such changes in habits simply show up as changes in the shape of the utility curve. Furthermore, this point of view is not empirically contentless. If it is to be useful, people must not make serious changes in their subjective value scale for a commodity as a result of relatively small changes in the amount of that commodity which they possess. This proposition is in principle testable, though testing it experimentally is difficult. However, Findings 4 and 5 of the probability preference experiments (see above) suggest that it may be correct.

Indifference Curves and a Property of NASEU Models

Loss of the addition theorem for SPs, even though compensated by a true zero for utility, is a very serious loss. Can a model of the type defined by Equation 1 mean much without an addition theorem for SPs? Yes. Equation 1 implies a strong decomposability property which has considerable significance both mathematically and empirically. The following discussion exhibits this property and derives two theorems, one of which can be empirically tested. These theorems have the additional advantage of being true for all plausible models defined by equations like Equation 1. This means that an empirical verification of them would be strong evidence for the usefulness of the class of maximization models with which this paper is concerned, while an empirical disproof would necessitate search for a quite different kind of theory of decision making (e.g., one based on variance preferences).

The following discussion deals exclusively with bets of the following form: Event E has probability p of occurring. If it occurs, S receives an amount of money x, which is greater than 0. If it does not occur, no money changes hands. Any such bet can be described as a point (x, p); that is, all such bets can be completely represented by a plane bounded at $p = 0$, $p = 1$, and $x = 0$. Now, consider the bets (x, p) and $[(x + e)p]$ where e is a small positive number. A great deal of evidence, plus common sense, suggests that most *S*s would prefer the second of these two bets to the first. Similarly, $[x, (p + f)]$ is preferable to (x, p) where f is another small positive number. If both of these assertions continue to be true as e and f approach zero for any values other than zero of x and p (a reasonable idealization), and if the direction of preference reverses for any negative values of e and f (another reasonable idealization), it follows that a set of indifference curves can be drawn in the xp plane. An indifference curve in this application is simply a function which defines a

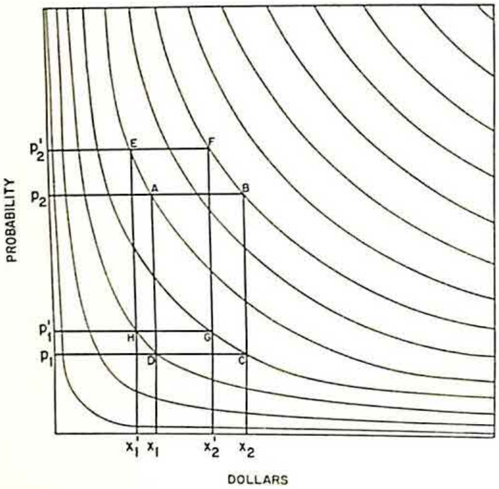


FIG. 1. Hypothetical indifference map among simple bets which have "no money changes hands" as one of the two possible outcomes. The rectangles are described in Theorems 4 and 5 of the text.

set of bets among all of which *S* is indifferent (see Coombs & Beardslee, 1954; Edwards, 1954c for details). A set of indifference curves is called an indifference map; Figure 1 shows such a map. The properties to be discussed below are general properties of a certain family of indifference curves and have applications to the use of indifference curves in economic theory; these applications will be discussed at the end of this section.

What does the SEU model have to say about such indifference curves? Since each curve identifies bets among all of which *S* is indifferent, all bets on a given curve must have the same SEU. In short, in utility and SP measures the equation of each indifference curve *c* must be:

$$\psi_c u_c = k_c \qquad [4]$$

Of course there is a different value of *k* for each curve.

But indifference curves are stated in physical, not subjective, units. What implications does Equation 4, and therefore SEU maximization models, have about the form of indifference curves stated in physical units? It is obvious that if any single indifference curve is selected, it is possible to find a transformation on the probability and money axes which will make it conform to Equation 4, which is of course the equation of a rectangular hyperbola. In fact, any set of two indifference curves could be made to conform by proper choice of transformations. But that uses up all the arbitrariness implied in Equation 1. To put it another way, the transformations necessary to make these two curves conform to Equation 4 completely specify the utility and subjective probability functions for this *S*. Now, if Equation 4 is to be correct, it must turn out that all other indifference curves in the *xp* plane must be turned into rectangular hyperbolas by the same utility and SP functions which turned the first two into rectangular hyperbolas. This is an exceedingly strong condition on the family of indifference curves. The following proof shows that an indifference map satisfies this requirement if and only if a quite specific relationship exists among the slopes of the indifference functions passing through any four points at the corners of a rectangle whose sides are parallel to the coordinate axes.

The theorem assumes that an indifference map is given each of whose functions is continuous, monotonic decreasing, and has continuous derivatives. The utility and SP transformations are assumed strictly monotonic increasing. They are also assumed single-valued, which means that the events displaying the OPs used must all have the same identifying characteristics if Theorems 4 and 5 are to be applicable. Two values of money, x_1 and x_2, and two values of probability, p_1 and p_2, are chosen at random (or systematically; it doesn't matter how they are chosen so long as both amounts of money are greater than zero and both probabilities lie between zero and one); they of course jointly define four bets. Let us call these four bets A, B, C, and D, as in Figure 1. From now on, S_A means the slope of the indifference curve which passes through Bet A.

Theorem 4: If any kind of SEU model, additive or otherwise, is correct, then $S_A S_C = S_B S_D$, and if this condition on the slopes of an indifference map is satisfied, and if at least one indifference function has a negative slope, then it follows that for these data a NASEU model or some more powerful SEU model is correct.[4]

The proof of Theorem 4 is in the mathematical appendix.

The property of all SEU maximization models, including NASEU models, which is established in Theorem 4 is theoretically elegant and in principle empirically testable. No such empirical test is likely to take place, however; differentiation of empirical curves requires more precision than is usually obtainable in experiments using human Ss. Fortunately, the microscopic property described in Theorem 4 has a macroscopic equivalent, which can be tested experimentally.

Look at Figure 1 again. Choose an amount of money x'_1, significantly smaller than x_1. Extend a vertical line from x'_1 to the indifference curve which passes through D; call the intersection point H. From H extend a vertical line until it intersects the indifference curve which passes through A, and a horizontal line until it intersects the indifference curve which passes through C; call the two points thus identified E and G. A horizontal line can be extended from E and a vertical one from G, intersecting in a point F and thus completing another rectangle. Will Point F lie on the indifference curve which passes through Point B? It turns out that it will if and only if the indifference map has the property expressed in Theorem 4. It is probably intuitively apparent that the property of Theorem 4 can be derived from this one; Theorem 4 is simply the limiting case as the difference between x_1 and x'_1 becomes infinitesimal. This property can be expressed formally as

Theorem 5: If any kind of SEU model, additive or otherwise, is correct, then if a second rectangle is constructed as in Figure 1 and three of its vertices are chosen to lie on the same indifference curves as three of the vertices of the original rectangle, then the fourth vertex of the new rectangle must lie on the same indifference curve as the fourth vertex of the original rectangle. Furthermore, if the vertices of the new rectangle are connected by straight lines to the vertices of the old rectangle, the slopes of the connecting lines will be related by the equation $S_{\overline{AE}}S_{\overline{CG}} = S_{\overline{BF}}S_{\overline{DH}}$, where $S_{\overline{AE}}$ symbolizes the slope of the line connecting Point A with Point E in Figure 1. If these conditions are satisfied, and if at least one indifference function has a negative slope, then it follows that for these data a NASEU model or some more powerful SEU model is correct.[5]

To prove Theorem 5, start by looking again at Equation 4. With notation taken from Figure 1, it says that:

$$g(x'_1)h(p'_1) = g(x_1)h(p_1) \qquad [5]$$

$$g(x'_1)h(p'_2) = g(x_1)h(p_2) \qquad [6]$$

$$g(x'_2)h(p'_1) = g(x_2)h(p_1) \qquad [7]$$

---

[4] This theorem and its proof were formulated by George Minty.

[5] This extension of Minty's theorem was formulated by L. J. Savage.

where $g(x_1)$ is the utility of x_1, $h(p_1)$ is the SP of p_1, and so on. Now multiply the left half of Equation 6 by the left half of Equation 7 and divide by the left half of Equation 5, and similarly for the right halves of these equations. The results are:

$$g(x'_2)h(p'_2) = g(x_2)h(p_2) \quad [8]$$

Equation 8 is what the first half of Theorem 5 asserts.

Now note that if a line is drawn from A to E and similarly for the other three pairs of vertices, each of these lines is a chord of an indifference curve. The slope of any of these lines is given by:

$$S_{ij} = \frac{p'_j - p_j}{x'_i - x_i} \quad [9]$$

where:

$$i, j = 1, 2$$

A little algebra will show that $S_{\overline{AE}}S_{\overline{CG}} = S_{\overline{BF}}S_{\overline{DH}}$, which is what the second half of Theorem 5 asserts. This result is a close macroscopic analogue of Theorem 4. In fact, Theorem 4 follows immediately from it if x'_1 is allowed to approach x_1, for then the four chord slopes approach the slopes of the indifference curves at the vertices of the original rectangle. This proof of Theorem 5 is only a proof of necessity; a proof of sufficiency very similar to that given in the Appendix for Theorem 4 exists, but will not be given here.

The property of NASEU (and other) models expressed in Theorem 5 can be tested by trying to perform the construction specified in the first half of Theorem 5. The relationship among slopes specified in the second half of Theorem 5 is tautologously true for all pairs of parallelograms with sides parallel to the coordinate axes, and so provides no additional information concerning the NASEU model. An experiment designed to test this property, and so to test this class of models, is now under way.

Although the foregoing discussion was expressed in terms of bets, it is actually a quite general property of certain kinds of indifference curves. The properties of Theorems 4 and 5 clearly apply to any indifference map whose curves have the form $f(x)g(y) = k$. But they also apply to any indifference map whose curves have the form $f(x) + g(y) = k$. (To prove this, simply replace the multiplications in Equations 5–7 by additions and the divisions by subtractions. The proof now proceeds as before, using additions instead of multiplications and subtractions instead of divisions up to but not including Equation 9. A similar modification of the proof of Theorem 4 is possible.) This additive form of the indifference curves corresponds to an indifference map concerned with two goods which are independent, rather than competing (like butter and margarine) or completing (like right and left shoes). Thus Theorems 4 and 5 provide an operational test of the hypothesis that two goods are independent of each other for a given indifference map—so far as I know, the only operational test so far suggested which can be applied directly to an indifference map. Of course this test is applicable only if the existence of at least interval scale utilities is assumed.

The Weighted SEU Model

Most of the argument of this paper has been devoted to the proposition that for any of the class of models considered here except NASEU models any function relating SP to OP may only be linear. The probability preference data indicate that it cannot be linear. Or do they?

An old familiar finding in psychophysics is that the form of any subjective scale depends on the methods used to determine it. The same proposition may be true for SP and utility scaling. The probability preference experiments were a rather indirect method inferring utility and SP properties from gambling choices. What results would more direct methods yield?

In the case of utility of money, it seems unlikely that direct methods of psychophysical scaling will be useful. The kind of question which would have to be asked would sound like "What amount of money is twice as valuable as two dollars to you?" It seems unlikely that any *S* would give an answer systematically different from "Four dollars," no matter what his utility function for money might look like. The difficulty is that money has an invincibly numerical character, and most people would probably respond to mathematical properties of these numbers rather than to any subjective values they might have. In any case, the evidence is continuing to accumulate that for small amounts of money utility is linearly related to dollar value—in which case psychophysical determination of utility scales is not very likely to be interesting. Of course psychophysical methods could be used to scale the utility of other objects which have a less numerical character, but it isn't obvious how such experiments would shed much light on general properties of utility functions, because in such experiments the concept of objective value is difficult to define.

In the case of SP, on the other hand, some psychophysical data exist. Shuford (1959) performed an experiment in which he showed *S*s square matrices made of 400 thumbtacks, some red and some green. The *S*s were required to estimate the relative frequency of each kind of thumbtack in the display after a brief look, and succeeded in estimating with remarkably good accuracy. Relative frequency in a matrix isn't exactly the same thing as OP, but is very closely related. Shuford's experiment suggested one in which I used these square matrices as the basis for probabilities in the probability preference experiment situation. I displayed two such matrices and associated a bet with each. The *S*s were required to estimate the relative frequency of each color in each of the two matrices, and then were required to choose one of the two bets. A random device (two 20-sided dice) chose one location on the matrix associated with the chosen bet, and the color of the thumbtack at that location determined the outcome of the bet. Preliminary results show that the estimates of relative frequency are excellent, just as in Shuford's experiments, but that the choices made among the bets are distorted. In short, at almost the same moment the *S*s say that SP is and is not equal to OP. Something peculiar is clearly going on.

My interpretation of this finding is that people perceive OPs correctly, but that they misuse them. They simply prefer to bet at some OPs rather than others, even though they perceive these OPs quite accurately. This hypothesis can be translated into mathematics by supposing that they attach to each event a weight w_i which expresses the relative desirability or undesirability of the probability displayed by that event. A model of this sort is only a minor variation on Equation 1:

$$\text{WSEU} = \sum_i (p_i w_i) u_i \qquad [10]$$

Mathematically, the WSEU model is exactly the same as the NASEU

This paper shows that if the subjective probabilities of a set of events of which some one must happen and not more than one can happen must add up to a specified constant (usually taken as one for convenience), then utility can be measured on an interval scale and subjective probability on a ratio scale. But if subjective probabilities do not have this additivity property, then both utility and subjective probability must be measured on a ratio scale.

If subjective probabilities do have the additivity property, if the decision model in which they are used is in principle applicable to any conceivable set of events, and if the number of different subjective probabilities which may occur in conjunction with a given objective probability is no more than denumerably infinite, then whenever objective probabilities are defined subjective probability must equal objective probability. This theorem and another mean that the concept of subjective probabilities which must add to one, and decision models based on that concept, cannot be very helpful in explaining decision making; a large amount of data indicates that subjective probability cannot always equal objective probability.

One alternative is to use a decision model in which subjective probabilities need not add up to any particular constant. Such a model is possible; its properties are explored. One property is that in such a model, utility must have a true zero point. The only reasonable zero point for utility is where you now are; this notion is not new to decision theory.

Any form of the expected utility maximization hypothesis (with additive or nonadditive subjective probabilities) implies a strong relationship among the slopes of indifference curves for simple bets. This relationship leads to an experimental test of the applicability of models of this class. If the result of this test should be negative, further theories about decision making in risky situations should probably include the concept of variance preferences.

REFERENCES

ALLAIS, M. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école americaine. *Econometrica*, 1953, **21**, 503–546. (a)

ALLAIS, M. La psychologie de l'homme rationnel devant le risque: La théorie et l'expérience. *J. Soc. Statist. Paris*, 1953, **94**, 47–73. (b)

COOMBS, C. H., & Beardslee, D. On decision-making under uncertainty. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954.

COOMBS, C. H., & PRUITT, D. G. *A study of decision making under risk. Willow Run Lab. Rep.* 1960, No. 2900–33–T. (Univer. Michigan.)

DAVIDSON, D., SUPPES, P., & SIEGEL, S. *Decision-making: An experimental approach.* Stanford. Stanford Univer. Press, 1957.

DVORETSKY, A., WALD, A., & WOLFOWITZ, J. Elimination of randomization in certain statistical decision procedures and zero-sum two-person games. *Ann. math. Statist.*, 1951, **22**, 1–21. (a)

DVORETSKY, A., WALD, A., & WOLFOWITZ, J. Relations among certain ranges of vector measures. *Pacific J. Math.*, 1951, **1**, 59–74. (b)

EDWARDS, W. Probability-preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

EDWARDS, W. Probability preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, **67**, 56–67. (a)

EDWARDS, W. The reliability of probability preferences. *Amer. J. Psychol.*, 1954, **67**, 68–95. (b)

EDWARDS, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380–417. (c)

EDWARDS, W. Variance preferences in gambling. *Amer. J. Psychol.*, 1954, **67**, 441–452. (d)

EDWARDS, W. The prediction of decisions among bets. *J. exp. Psychol.*, 1955, **50**, 201–214.

EDWARDS, W. *Subjective probability in decision theories. Willow Run Lab. Rep.*, 1959, No. 2144–361–T. (Univer. Michigan.)

EDWARDS, W. Measurement of utility and subjective probability. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960.

EDWARDS, W. Behavioral decision theory. *Annu. Rev. Psychol.*, 1961, **12**, 473–498.

FRIEDMAN, M., & SAVAGE, L. J. The utility analysis of choices involving risk. *J. pol. Econ.*, 1948, **56**, 279–304. (Reprinted with minor changes in G. J. Stigler & K. E. Boulding [Eds.], *Readings in price theory.* Chicago: Richard D. Irwin, 1952.)

FRIEDMAN, M., & SAVAGE, L. J. The expected-utility hypothesis and the measurability of utility. *J. pol. Econ.*, 1952, **60**, 463–475.

HALMOS, P. R. *Measure theory.* New York: Van Nostrand, 1950.

LUCE, R. D. On the possible psychophysical laws. *Psychol. Rev.*, 1959, **66**, 81–95.

LYAPOUNOV, A. Sur les fonctions-vecteurs complètement additives. *Izv. Akad. Nauk SSSR Ser. Math.*, 1940, **4**, 465–478.

MARKOWITZ, H. The utility of wealth. *J. pol. Econ.*, 1952, **60**, 151–158.

MOSTELLER, F., & NOGEE, P. An experimental measurement of utility. *J. pol. Econ.*, 1951, **59**, 371–404.

SAVAGE, L. J. *The foundations of statistics.* New York: Wiley, 1954.

SHUFORD, E. H. *A comparison of subjective probabilities for elementary and compound events.* Chapel Hill, N. C.: University of North Carolina, Psychometric Laboratory, 1959. (Report No. 20)

VON NEUMANN, J., & MORGENSTERN, O *Theory of games and economic behavior.* (2nd ed.) Princeton: Princeton Univer. Press, 1947.

(Received August 23, 1960)

APPENDIX[A1]

This appendix contains precise statements of the assumptions and proofs of Theorems 2, 3, and 4 from the body of the article. The proofs of Theorems 1 and 5 are contained in the text.

Theorems 2 and 3 assume the existence of a probability space S with events $A, B, C, \cdots$, having objective probabilities $p(A), p(B), p(C), \cdots$; the measure p has the conventional properties. Associated with each Event A is an element $r(A)$ of a space R; $r(A)$ is called the identifying characteristic of Event A. Two or more events (indeed, large classes of events in the models discussed in this paper) may have the same identifying characteristics. A function $\Gamma(r, p)$ is defined for each r and p such that for some Event A, $r(A) = r$ and $p(A) = p$. The subjective probability $\psi(A)$ of the Event A is then defined by the equation:

$$\psi(A) = \Gamma[r(A), p(A)]$$

Several assumptions, mostly of technical importance only, are necessary. They are stated here with little discussion; the important ones are discussed in the text.

Axiom 1: p is completely additive.

Axiom 2: S is nonatomic with respect to p. This means that if $p(A) > 0$, then there exists some Event B contained in A (a subset of A) such that $0 < p(B) < p(A)$.

Axiom 3: For each r for which $\Gamma(r, 0)$ is defined, $\Gamma(r, 0) = 0$, and $\lim_{p \to 0} \Gamma(r, p) = 0$ uniformly in r. That is, given $\epsilon > 0$ there exists a $\delta > 0$ such that $\Gamma(r, p) < \epsilon$ whenever $p < \delta$ and $\Gamma(r, p)$ is defined.

Axiom 4: $\psi(A)$ is a probability measure. Specifically: (*a*) For $A \subset S$, $0 \leqq \psi(A) \leqq 1$, (*b*) $\psi(S) = 1$ and $\psi(\bar{S}) = 0$, and (*c*) If $A, B, \subset S$ and $A \cap B = 0$, then $\psi(A \cup B) = \psi(A) + \psi(B)$.

There is some redundancy in the statement of Axiom 4; it was included for clarity and to permit Theorem 4 to be proved without the use of Part *c* of the axiom.

Some lemmas prepare the way for the proof of Theorems 2 and 3.

Lemma 1: The probability measure ψ is completely continuous with respect to p, and completely additive.

From Axiom 3, it follows immediately that to any number $\epsilon > 0$ there corresponds a number $\delta > 0$ such that $\psi(A) < \epsilon$ for any Event A such that $p(A) < \delta$. This is what complete continuity means.

[A1] See body of article for credit footnotes identifying the formulators of each theorem.

Now let $A_1, A_2, A_3, \cdots$ be a countable sequence of disjoint events such that $\bigcup_{n=1}^{\infty} A_n = A$. Since p is completely additive, we have $p(A) = \sum_{n=1}^{\infty} p(A_n)$ and (setting $B_n = \bigcup_{k=n}^{\infty} A_k$) $\lim_{n\to\infty} p(B_n) = 0$. It follows from the foregoing discussion that $\lim_{n\to\infty} \psi(B_n) = 0$. Since $p(A)$ exists, so does $\psi(A)$, and since it follows from Axiom 4 that $\psi(A) = \sum_{k=1}^{n-1} \psi(A_k) + \psi(B_n)$ for all n, we conclude that $\psi(A) = \sum_{n=1}^{\infty} \psi(A_n)$, and thus that ψ is completely additive

Lemma 2: S is nonatomic with respect to ψ.

Proof: Lemma 1 makes it possible to apply the Radon-Nikodym theorem (see for example Halmos, 1950, p. 128) to infer that ψ has a probability density with respect to p. If A is any event with $\psi(A) > 0$, it must include a Subevent A' such that $p(A') > 0$ and the density of ψ with respect to p is positive on A'. Since S is nonatomic with respect to p by Axiom 2, there is an event $B \subset A'$ such that $0 < p(B) < p(A')$; consequently also $0 < (B) < \psi(A') \leqq \psi(A)$. This proves the lemma.

Lemma 3: Let f be a (Borel-measurable) function defined on S such that $0 \leqq f(w) \leqq 1$ for w in S, and let Ω_1 and Ω_2 be completely additive nonatomic probability measures. Then there exists an Event E such that:

$$\int_S f(w)\,d\Omega_1(w) = \Omega_1(E),$$

$$\int_S f(w)\,d\Omega_2(w) = \Omega_2(E)$$

This result is a specialization of one established by Dvoretsky, Wald, and Wolfowitz (1951a, 1951b), representing an extension of an earlier theorem of Lyapounov (1940).

It is convenient at this point to interpret the objective and subjective probabilities of an Event A as the rectangular coordinates of a point $[p(A), \psi(A)]$ in a Euclidean plane. Then if p and ψ are not identical measures there must exist an Event H for which $\psi(H) \neq p(H)$, and the points $(0, 0)$, $[p(H), \psi(H)]$, $[p(\bar{H}), \psi(\bar{H})]$, and $(1, 1)$ (where $\bar{H}$ is the complementary event to H) form a parallelogram. The following lemma asserts that there is at least one event corresponding to every point within the parallelogram.

Lemma 4: Let H be any event, and let a and b be any two numbers such that $0 \leqq a \leqq 1$ and $0 \leqq b \leqq 1$. Then there exists an Event C such that $p(C) = ap(H) + b[1 - p(H)]$ and $\psi(C) = a\psi(H) + b[1 - \psi(H)]$.

Proof: Let g be the characteristic function of H (that is, let $g(w)$ be 1 for w in H, and 0 otherwise). Then by Axioms 1 and 2 and Lemmas 1 and 2 the hypotheses of Lemma 3 are satisfied with $\Omega_1 = p$, $\Omega_2 = \psi$, and $f(w) = ag(w) + b[1 - g(w)]$. The Event C is then the event whose existence is asserted by Lemma 3.

Now it is possible to prove

Theorem 2: If there is an Event H for which $\psi(H) \neq p(H)$, then for any value of p within the limits $0 < P < 1$ there exist Events A and B satisfying the relations $p(A) = p$, $p(B) < p(A)$, $\psi(B) > \psi(A) + P(1 - P)|\psi(H) - p(H)|$.

Proof: It is sufficient to consider the case $p(H) \leqq 1/2$, $P \leqq 1/2$, $\psi(H) - p(H) = h > 0$; parallel arguments establish the result for other cases.

Suppose first that $0 < P \leqq p(H)$. By Lemma 4 (with $a = 0$ and $b = P/[1 - p(H)]$) there exists an Event A such that $p(A) = P$, while $\psi(A) = P[1 - \psi(H)]/[1 - p(H)]$. Similarly, taking $a = P/p(H)$, $b = 0$ yields the conclusion that there exists an Event B' with $p(B') = P$, $\psi(B') = P\psi(H)/p(H)$.

Subtraction yields the inequality:

$$\psi(B') - \psi(A) = \frac{hP}{p(H)[1 - p(H)]} > hP(1 - P)$$

Next suppose $p(H) < p \leqq 1/2$. The Event A may be selected as in the foregoing case, while an appropriate Event B' corresponds to $a = 1$, $b = [P - p(H)]/[1 - p(H)]$. Thus in this case also:

$$\psi(B') - \psi(A) = \psi(H) + \frac{P - p(H)}{1 - p(H)}[1 - \psi(H)] - P\frac{1 - \psi(H)}{1 - p(H)} = \frac{h}{1 - p(H)} > hP(1 - P)$$

Now let the values of a and b corresponding to the Event B' be multiplied by the factor:

$$k = 1/2\left[1 + \frac{hP(1 - P) + \psi(A)}{\psi(B')}\right]$$

and designate an associated event by B. Since $k < 1$, it follows that:

$$p(B) = kp(B') < P = p(A)$$

while

$$\psi(B) = k\psi(B') = 1/2[\psi(B') + hP(1 - P) + \psi(A)] > hP(1 - P) + \psi(A)$$

Thus the Events A and B have the desired properties and the proof is complete.

Theorem 2 could be strengthened, in that the expression $hP(1 - P)$ could be replaced by any continuous function $G(P)$ satisfying the inequalities:

$$G(P) < \frac{hP}{p(H)[1 - p(H)]}, \quad 0 < P \leqq \min[p(H), 1 - p(H)]$$

$$G(P) < \frac{h}{\max[p(H), 1 - p(H)]}$$

$$\min[p(H), 1 - p(H)] \leqq P \leqq \max[p(H), 1 - p(H)]$$

$$G(P) < \frac{h(1 - P)}{p(H)[1 - p(H)]}, \quad \max[p(H), 1 - p(H)] \leqq P < 1$$

The proof is changed only by the replacement of $hP(1 - P)$, where it appears, by $G(P)$.

Lemma 4 also makes it possible to draw a strong conclusion about the cardinality of the set of identifying characteristics.

Theorem 3: If there exists an Event H for which $\psi(H) \neq p(H)$, the set R of identifying characteristics is nondenumerably infinite.

Proof: Let the hypothesis of the theorem be satisfied for some H. Suppose for convenience that:

$$p(H) \leqq 1/2 \leqq 1 - p(H)$$

Then

$$ap(H) + b[1 - p(H)] = 1/2$$

if

$$b = [1/2 - ap(H)]/[1 - p(H)]$$

Since the foregoing relations imply that $0 \leqq b \leqq 1$ if $0 \leqq a \leqq 1$, Lemma 4 applies, and for each value of a in the indicated range there exists an Event C_a such that $p(C_a) = 1/2$, while:

$$\psi(C_a) = \frac{1/2[1 - \psi(H)] + a[\psi(H) - p(H)]}{1 - p(H)}$$

Thus distinct values of a are associated with distinct values of $\psi(C_a)$. Since $p(C_a)$ is the same for all a, the identifying characteristics $r(C_a)$ must likewise be distinct for distinct a's. Thus a subset of R can be put into one-to-one correspondence with the unit interval, and the theorem is established.

The general mathematical setup underlying Theorem 4 is explained in the body of the paper; it is not altogether the same as that underlying Theorems 2 and 3. Theorem 4 is concerned with bets of the form "with probability p you win \$$x$; with probability $1 - p$ no money changes hands." Specifically, it assumes the existence of indifference curves in a plane of such bets, and inquires about the restrictions which a NASEU model puts on such curves.

Axioms 1 and 2 of this Appendix still apply, Axiom 3 is irrelevant, and Parts *a* and *b* of Axiom 4 are assumed (but *not* Part *c*). The following additional axioms are assumed.

Axiom 5: For the particular situations in which Theorems 4 and 5 are to be tested, $\psi(p)$ is a continuous, differentiable, monotonically increasing function of p.

Axiom 6: For the particular situations in which Theorems 4 and 5 are to be tested, $u(x)$ is a continuous, differentiable, monotonically increasing function of x.

Axiom 7: The model asserted by Equation 1 of the text applies. That is, in this situation Bet A is preferred to Bet B if and only if:

$$\psi(p_A)u(x_A) > \psi(p_B)u(x_B)$$

Now we perform the construction described in the paper and illustrated in Figure 1. If S_A is the slope of the indifference curve passing through Bet A, we can prove

Theorem 4: $S_A S_C = S_B S_D$.

First comes a proof that if the indifference curves can be transformed into a form fitting Equation 4 of the paper, then the first half of Theorem 4 must hold. If Equation 4 can be written for any indifference curve, there must exist two functions g and h such that the equation for each indifference curve can be written:

$$g(x)h(p) = k_c \qquad [A1]$$

Differentiating Equation 11 gives:

$$g'(x)h(p)dx + g(x)h'(p)dp = 0$$

$$\frac{dp}{dx} = -\frac{g'(x)}{g(x)}\cdot\frac{h(p)}{h'(p)} \qquad [A2]$$

$$\frac{dp}{dx} = \Phi(x)\theta(p) = S(x, p) \qquad [A3]$$

which simply says that the slope of the indifference curve at any point may be written as a product of a function of x along and a function of p alone. Now it is tautologous that:

$$\frac{\Phi(x_1)\theta(p_2)}{\Phi(x_1)\theta(p_1)} = \frac{\Phi(x_2)\theta(p_2)}{\Phi(x_2)\theta(p_1)} \qquad [A4]$$

Substitution from Equation A3 reduces Equation A4 to:

$$\frac{S(x_1, p_2)}{S(x_1, p_1)} = \frac{S(x_2, p_2)}{S(x_2, p_1)} \qquad [A5]$$

But by definition (x_1, p_2) defines Bet A, (x_2, p_2) defines Bet B, (x_2, p_1) defines Bet C, and (x_1, p_1) defines Bet D. So Equation A5 may be rewritten:

$$S_A S_C = S_B S_D \qquad [A6]$$

The proof so far has shown that if Equation 4 is correct then Equation A6 follows. In short, Equation A6 is necessary for the correctness of Equation 4, and so of Axiom 7. It remains to show that it is also sufficient. The statement of Theorem 4 given in the text of the paper specifies that Equation A6 is sufficient as well as necessary for the correctness of Equation 4 only if at least one indifference curve has a negative slope; this restriction is necessary to rule out transformations on p and x which are monotonically decreasing rather than monotonically increasing. To prove sufficiency given at least one indifference curve with a negative slope, it is necessary to exhibit transformations which actually fit Equation 4.

To start, choose any point (x_0, p_0), $x_0 > 0$, $0 < p_0 < 1$. Now let

$$u = \exp - \int_{x_0}^{x} S(x, p_0)dx \qquad [A7]$$

$$\psi = \exp S(x_0, p_0) \int_{p_0}^{p} \frac{1}{S(x_0, p)}\,dp \qquad [A8]$$

Now it is necessary to prove that these transformations fit Equation 4 if they fit Equation A6.

First, differentiate Equations A7 and A8:

$$d\psi = \left[\exp S(x_0, p_0) \int_{p_0}^{p} \frac{1}{S(x_0, p)}\, dp\right] \times \left[S(x_0, p_0)\, \frac{1}{S(x_0, p)}\, dp\right]$$

$$du = \left[\exp - \int_{x_0}^{x} S(x, p_0) dx\right] \times [-S(x, p_0) dx]$$

So

$$\frac{d\psi}{du} = -\frac{\psi}{u}\cdot\frac{S(x_0, p_0)}{S(x, p_0) S(x_0, p)}\,\frac{dp}{dx} \quad [\text{A9}]$$

Substituting into A9 from A6:

$$\frac{d\psi}{du} = -\frac{\psi}{u}\cdot\frac{S(x_0, p_0)}{S(x, p) S(x_0, p_0)}\,\frac{dp}{dx}$$

$$\frac{d\psi}{du} = -\frac{\psi}{u}\cdot\frac{1}{S(x, p)}\cdot\frac{dp}{dx}$$

But $\frac{dp}{dx} = S(x, p)$ by definition. Therefore, $\frac{d\psi}{du} = -\frac{\psi}{u}$ which is the differential equation for the family of rectangular hyperbolas.

Psychological Review
1962, Vol. 69, No. 2, 136–148

DECISION MAKING:

OBJECTIVE MEASURES OF SUBJECTIVE PROBABILITY AND UTILITY[1]

GORDON M. BECKER[2]

General Dynamics Corporation, Groton, Connecticut

This paper describes several methods for measuring subjective probability. Some of the methods require a priori measures of utility, others measure both utility and subjective probability simultaneously. The measures, derived from choice behavior, do not necessarily reflect the subject's (*S*'s) conscious opinions or feelings. The subjective probability estimates may be useful as measures of confidence, bias, prejudice, or expectancy.

The basis for the methods is a modification of Edwards (1955) Subjective Expected Utility (SEU) model. It is assumed that:[3]

1. A person confronted with a choice between two or more alternatives selects the alternative that appears, to him, to offer the highest average satisfaction (SEU). The SEU of an alternative, A, is defined as:

$$\mathrm{SEU}(A) = \sum_{i=1}^{I} \psi(A_i)u(A_i)$$

where:

A_i is the ith possible outcome of course of action A; $i = 1, 2, \cdots, I$.
$\psi(A_i)$ is S's subjective probability that Outcome A_i will occur.
$u(A_i)$ is the satisfaction (utility) S expects to derive if Outcome A_i occurs when S takes course of action A.

2. The sum of the subjective probabilities of a set of mutually exclusive and exhaustive events equals unity,[4] i.e.:

$$\sum_{i=1}^{I} \psi(A_i) = 1$$

3. Utility and subjective probability do not interact.

4. The satisfaction (utility) S associates with an event is measurable.

5. The measures of S's utility reflect the utilities operating during the subjective probability measurement session, i.e., the utilities have not changed during the period intervening between the utility and the subjective probability measurement sessions.

In addition to the above, the assumptions inherent in the method used to measure S's utilities must be added. If the utility measures are based upon the assumption that utility differences of all magnitudes are discernable (Coombs & Komorita, 1958; Edwards, 1955; Luce & Raiffa, 1957; Savage, 1954; Siegel, 1956), i.e., that any difference in SEU leads to 100% preference for the alternative with the higher SEU, this assumption becomes part of the present model.

---

[1] The major ideas discussed in the first half of this paper were presented at the Thirteenth Annual Meeting of the Eastern Psychological Association, Atlantic City, April 3, 1959. This investigation was supported in part by Research Grant M-2348(RI) from the National Institute of Mental Health, United States Public Health Service, in part by the General Dynamics Corporation, and in part by Itek Corporation.

[2] Now located at TEMPO, General Electric Company.

[3] The author is indebted to Ward Edwards for his help in making these assumptions explicit, for encouraging the author to pursue these ideas, and for suggesting that the author consider the effects of various degrees of information about utility upon the determination of subjective probability.

[4] This is a modification of Edwards SEU model in that he did not require that the sum of subjective probabilities equal unity.

However, if the utility measures are based upon the assumption that preferences are stochastic when utility differences are small (Coombs, 1959; Davidson & Marschak, 1959; Davidson, Suppes, & Siegel, 1957; Luce, 1959; Suppes & Walsh, 1959), then the stochastic assumptions become part of the present model.

Figure 1 shows the SEUs of two alternatives Y_1 and Y_2 as functions of subjective probability. The figure illustrates a situation in which you, as *S*, must choose between going to the airport (Y_1) or going to the Green Room (Y_2). In this example you want to meet Bill, who will be either (B_1) at the airport, or (B_2) at the Green Room. The satisfactions (utilities) you expect to derive from the four possible outcomes of this situation are given by ϕ_{11}, ϕ_{12}, ϕ_{21}, and ϕ_{22}.

The model predicts that you will choose Y_1 (the airport) whenever your subjective probability that Bill is at the airport, $\psi(B_1)$, is greater than X_1. Note that X_1 is the subjective probability at which the two alternatives have equal SEU and intersect. (If a stochastic model is used to determine utilities, X_1 will be a range rather than a point.) Note also that the sum of the subjective probabilities, $\psi(B_1) + \psi(B_2)$, is unity, indicating that you are sure Bill will be either at the airport or the Green Room, not at both, and not somewhere else.

MEASURING SUBJECT PROBABILITY

The subjective probability method proposed in this paper is based upon the reverse of the above interpretation, i.e., instead of predicting the alternative *S* chooses, the method estimates *S*'s subjective probability at the time of choice. From Figure 1, it is evident that when *S* chooses Alternative Y_1, his subjective probability $\psi(B_1)$, must be between X_1 and unity, whereas when *S* chooses Y_2, $\psi(B_1)$ must be between X_1 and 0. Knowledge of both the utilities and the choice permits one to place upper and lower bounds on $\psi(B_1)$.

YOU CHOOSE Y	BILL CHOOSES B	
	AIRPORT (B1)	GREEN Rm (B2)
AIRPORT (Y1)	$\phi(11)$	$\phi(12)$
GREEN Rm (Y2)	$\phi(21)$	$\phi(22)$

WHERE: $\phi(22) > \phi(11) > \phi(21) > \phi(12)$

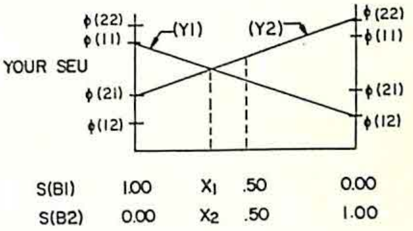


FIG. 1. Hypothetical problem illustrating SEU model.

The range between the upper and lower bounds can be decreased by having *S* choose from among several alternatives. In general, if *N* different utilities are known, it is possible to construct N^2 different two-outcome alternatives.

Let $(a_i, p; a_j)$ represent an alternative, *A*, that pays a_i when *p* occurs and a_j otherwise. If $u(a_1) > u(a_2) > \cdots > u(a_n)$ it follows that $(a_1, p; a_1)$ is preferred to all other alternatives at all probabilities except $\psi(p) = 1$ at which point all alternatives of the form $(a_1, p; x)$ are equally preferred and except at the point $\psi(p) = 0$ when all alternatives of the form $(x, p; a_1)$ are equally preferred. Alternative $(a_1, p; a_2)$ is preferred to all alternatives of the form $(a_1, p; x)$ when $u(x) < u(a_2)$ and is preferred to all alternatives of the form $(x, p; a_2)$ when $u(x) < u(a_1)$. However, there are $N - 1$ alternatives of the form $(x, p; a_1)$ which are preferred to $(a_1, p; a_2)$ when $\psi(p)$ is small, but are not preferred when $\psi(p)$ is large.

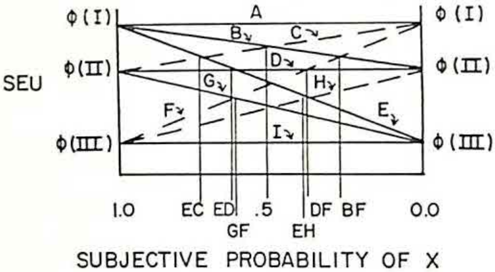


FIG. 2. Example illustrating relationship between SEU, subjective probability, and alternative choices.

Hence, these $N-1$ alternatives intersect $(a_1, p; a_2)$ at various $\psi(p)$. In general, alternatives of the form $(a_i, p; a_j)$ intersect $N-i$ alternatives of the form $(x, p; y)$, when $u(x) < u(a_i)$ and $u(y) > u(a_j)$. Figure 2 shows the nine two-outcome alternatives that can be formed from three distinct utilities and the nine subjective probability intersection points that result. In general, N distinct utilities will result in N^2 two-outcome alternatives and $N^2(N-1)^2/4$ intersections.

Let the subjective probability at the point at which Alternative A intersects Alternative B be $\psi(A, B)$—where $A=(a, p; a')$ and $B=(b, p; b')$. Since the SEUs of A and B are equal at the intersection point, it follows that:

$$\psi(p)u(a) + [1-\psi(p)]u(a') = \psi(p)u(b) + [1-\psi(p)]u(b')$$

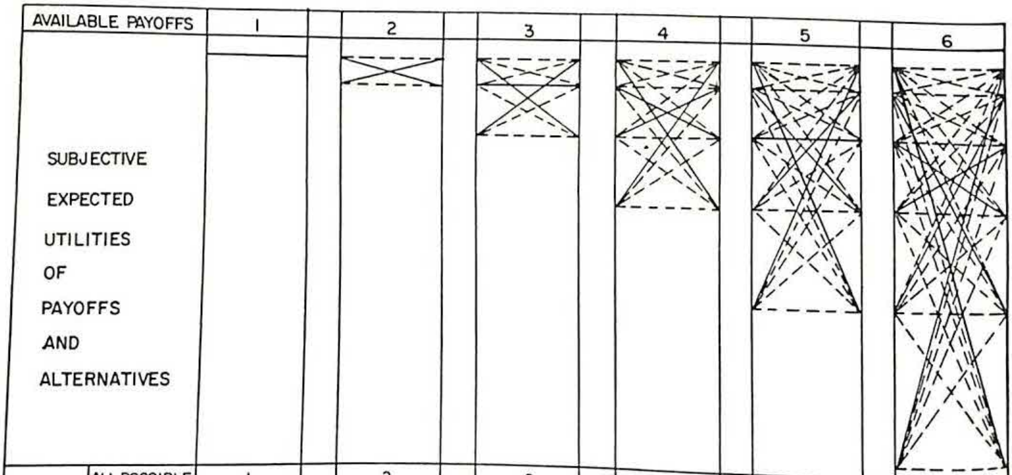
and:

$$\psi(A, B) = \frac{u(a') - u(b')}{u(b) - u(b') - u(a) + u(a')}$$

It is evident that $\psi(A, B) = \psi(B, A)$, i.e., that reversing the order of the alternatives does not change the subjective probability at the intersection point.

However, reversing the payoffs (utilities) within an alternative will change the crossover point. Let Alternative A' be $(a', p; a)$ and let Alternative B' be $(b', p; b)$. It follows that:

$$\psi(A', B') = \frac{u(a) - u(b)}{u(b') - u(b) - u(a') + u(a)}$$



	AVAILABLE PAYOFFS	1	2	3	4	5	6
NUMBER OF PREFERENCE RANKS REQUIRED OF SUBJECT	ALL POSSIBLE	1	2	8	32	92	210
	FIVE	1	2	4	6	10	14
	FOUR	1	2	4	6	10	14
	THREE	1	2	4	6	10	12
	TWO	1	2	4	6	8	10
	ONE	1	2	3	4	5	6
		MAXIMUM NUMBER OF OBTAINABLE SUBJECTIVE PROBABILITY INTERVALS					

FIG. 3. The number of obtainable subjective probability intervals as a function of the number of required choices and the number of available payoffs.

Comparing $\psi(A', B')$ with $\psi(A, B)$ it will be seen that $\psi(A', B') = 1 - \psi(A, B)$.

It is also evident, from the above equations, that $\psi(A, A') = 1/2$, i.e., reversing the payoffs within an alternative will result in the reversed alternative intersecting the original at $\psi(p) = 1/2$. In general, $N(N - 1)/2$ of the intersections of the two-outcome alternatives formed from N unique utilities will intersect at $\psi(p) = 1/2$. One need know only the ranks of the two utilities used in these alternatives to determine which alternative has the higher SEU at subjective probabilities greater (or less) than 1/2.

The maximum number of unique intersections of two-outcome alternatives that can be achieved from N unique utilities is therefore:

$$\frac{(N^2 - 1)(N^2 - 2N) + 4}{4}$$

The maximum number of subjective probability intervals that can be determined from N unique utility points is one more than the number of unique intersections, as shown in Figure 3.

Utilities Measured on Interval or Ratio Scales

One method for estimating the subjective probability operating at the time of *S*'s choice is to have *S* select one of several alternatives. For example, if *S* must choose between Alternatives *E*, *D*, and *F* shown in Figure 2 and he chooses Alternative *E*, his subjective probability, according to the model, must be between *ED* and unity. Whenever *S* selects Alternative *F*, his subjective probability must be between *DF* and 0; and when *S* selects Alternative *D*, his subjective probability must be between *ED* and *DF*. If the utilities, $\phi(\text{I})$, $\phi(\text{II})$, and $\phi(\text{III})$ are measured on a ratio or interval scale, the numerical values of these probability bounds can be determined from the equations for the intersections of the alternatives.

The range between the upper and lower bounds can be reduced by having *S* again choose between alternatives. If *S*'s original choice is *E*, Alternatives *E* and *C* might be presented. Given the initial preference *E*, if *S* now selects *E* over *C* his subjective probability must be between unity and *EC*, whereas a choice for *C* at this point would indicate a subjective probability between *EC* and *ED*. Similarly, Alternatives *E* and *F* might be presented if the original choice were *D*, and Alternatives *B* and *F* might be presented if the original choice were *F*. Given three utilities, it is possible, from two responses, to estimate which of six intervals contains *S*'s subjective probability. A third choice would permit the experimenter to determine which of eight subjective probability ranges contains *S*'s subjective probability.

Another technique for obtaining increased precision requires only a single response. *S* ranks his preferences for many alternatives instead of simply selecting the one he prefers most. Figure 3 shows the maximum number of intervals obtainable with from one to six utility points. Notice that the number of intervals depends upon the depth of the ranking required of *S* and the number of utility points used.

Utilities Measured on an Ordinal Scale

If only the ranks of the utilities are known, at least $2N - 2$ ranked probability intervals can be obtained. As indicated earlier, knowledge of the ranks of two utilities is sufficient information to enable the experimenter to construct a pair of alternatives such that *S*'s choice indicates whether his subjective probability is greater than, equal to, or less than 1/2. Alternatives B and C, and E and F, and G and H (Figure 2) are examples of such alternatives.

In order to obtain $2N - 2$ ranked probability from N different utilities when only the ranks of the utilities are known, construct a set of $2N - 3$ two-outcome alternatives. Let one alternative, AN, consist of winning the highest utility, a, when Event p occurs, and the lowest utility, n, otherwise, i.e., $AN = (a, p; n)$. From each of the other $N - 2$ utilities construct two alternatives. Let one alternative from each utility be a sure thing, and let the other alternative consist of winning the next lower utility when p does not occur, e.g., from Utility i construct Alternatives II and IJ where $II = (i, p; i)$ and $IJ = (i, p; j)$ and where the rank of $u(i)$ is one higher than the rank of $u(j)$. It is evident that Alternative II intersects AN at a higher subjective probability than the intersection of AN and IJ, i.e., $\psi(AN, II) > \psi(AN, IJ)$. Moreover, it is evident that $\psi(AN, BB) > \psi(AN, BC) > \psi(AN, CC) > \cdots > \psi(AN, MM)$ where $u(a) > u(b) > \cdots > u(m) > u(n)$. Two additional alternatives can be formed to complete the set of $2N - 2$ alternatives. One of these alternatives, $BA = (b, p; a)$, will intersect AN at the highest subjective probability, and the other alternative, $NM = (n, p; m)$ will intersect AN at the lowest subjective probability. Therefore, the $2N - 3$ alternatives formed in this manner intersect AN at $2N - 3$ different subjective probabilities. The ranks of the subjective probabilities at the intersection points are strictly determined by the ranks of the utilities. Hence, from N different utilities it is possible to construct a set of alternatives such that $2N - 2$ subjective probability intervals can be ranked from a knowledge of only the utility ranks.

A second technique for measuring subjective probability when only the utility ranks are known makes it possible to measure both utility and subjective probability simultaneously. Let the N utilities be ranked as before: $u(a) > u(b) > \cdots > u(n)$. Construct the following N two-outcome alternatives: $AN = (a, p; n)$; $BM = (b, p; m)$; $CL = (c, p; l)$; $\cdots$; $NA = (n, p; a)$. These N alternatives intersect, at most, at $(N^2-2N+2)/2$ distinct probability points to form $(N^2-2N+4)/2$ probability intervals. Consider the case of four utilities, ranked a highest and d lowest. Let the four two-outcome alternatives, constructed in accordance with the above plan, be: $A = (a, p; d)$; $B = (b, p; c)$; $C = (c, p; b)$; and $D = (d, p; a)$. The difference between Utilities a and b, i.e., $u(a) - u(b)$, must be less than, equal to, or greater than the difference between Utilities c and d. But when $u(a) - u(b) < u(c) - u(d)$, the ranks of the SEUs of the Alternatives A, B, C, and D are:

$D>C>B>A$ when $0 \le \psi(p) < \psi(C, D)$
$D=C>B>A$ when $\psi(p) = \psi(C, D)$
$C>D>B>A$ when $\psi(C, D) < \psi(p) < \psi(B, D)$
$C>D=B>A$ when $\psi(p) = \psi(B, D)$
$C>B>D>A$ when $\psi(B, D) < \psi(p) < 1/2$
$C=B>D=A$ when $\psi(p) = 1/2$
$B>C>A>D$ when $1/2 < \psi(p) < \psi(A, C)$
$B>C=A>D$ when $\psi(p) = \psi(A, C)$
$B>A>C>D$ when $\psi(A, C) < \psi(p) < \psi(A, B)$
$A=B>C>D$ when $\psi(p) = \psi(A, B)$
$A>B>C>D$ when $\psi(A, B) < \psi(p) \le 1$

Note that only 8 of the 24 possible rankings of A, B, C, D occur when $u(a) - u(b) < u(c) - u(d)$.

When $u(a) - u(b) = u(c) - u(d)$, the SEU ranks are:

$D>C>B>A$ when $0 \le \psi(p) < 1/2$
$A=B=C=D$ when $\psi(p) = 1/2$
$A>B>C>D$ when $1/2 < \psi(p) \le 1$

Note that all 24 rankings of A, B, C, D are possible when $u(a) - u(b) = u(c)$

$- u(d)$, but only two rankings occur when $\psi(p)$ is different from 1/2.

When $u(a) - u(b) > u(c) - u(d)$, the SEU ranks are:

$D > C > B > A$ when $0 \leq \psi(p) < \psi(A, B)$
$D > C > B = A$ when $\psi(p) = \psi(A, B)$
$D > C > A > B$ when $\psi(A, B) < \psi(p) < \psi(A, C)$
$D > C = A > B$ when $\psi(p) = \psi(A, C)$
$D > A > C > B$ when $\psi(A, C) < \psi(p) < 1/2$
$D = A > C = B$ when $\psi(p) = 1/2$
$A > D > B > C$ when $1/2 < \psi(p) < \psi(B, D)$
$A > B = D > C$ when $\psi(p) = \psi(B, D)$
$A > B > D > C$ when $\psi(B, D) < \psi(p) < \psi(C, D)$
$A > B > C = D$ when $\psi(p) = \psi(C, D)$
$A > B > C > D$ when $\psi(C, D) < \psi(p) \leq 1$

Note that only 8 of the possible 24 rankings of A, B, C, D occur when $u(a) - u(b) > u(c) - u(d)$.

From the above, it is evident that, when $\psi(p)$ is different from 1/2, six rankings of the alternatives occur only when $u(a) - u(b) > u(c) - u(d)$ and six other rankings occur only when $u(a) - u(b) < u(c) - u(d)$. Thus if S assigns any of these 12 distinctive rankings when $\psi(p)$ is not 1/2, the experimenter can immediately rank the subjective probability of p operating at that time and, at the same time, determine whether $u(a) - u(b)$ is greater or less than $u(c) - u(d)$. If $u(a) - u(b)$ is greater (less) than $u(c) - u(d)$, 10 of the 24 rankings will never occur, and if the utility differences are equal then 22 of the 24 rankings will never occur when $\psi(p)$ is different from 1/2. Therefore when subjective probabilities different from one-half are used, the experimenter can immediately determine from S rankings, whether $u(a) - u(b)$ is greater than or less than $u(c) - u(d)$, can immediately order the six subjective probability intervals obtainable from the four utilities, and can simultaneously determine which of the six intervals contains the $\psi(p)$ operative at the time of S's ranking.

Thus, given only the ranks of the utilities, it is possible to order the utility differences and the subjective probability intervals from S's rankings of the alternatives.

However, the above method does not allow for stochastic choice. If S ranks the alternatives consistent only with $u(a) - u(b) > u(c) - u(d)$ at one time, and consistent only with $u(a) - u(b) < u(c) - u(d)$ at another time, the method will lead to contradictory results. If people show such inconsistencies in their rankings it will be necessary to modify the method by extending it to include stochastic rankings of alternatives.

PRELIMINARY TEST OF THE GENERAL METHOD

Two preliminary experiments were conducted to explore whether or not Ss vary their selection of alternatives in accordance with the difficulty of a discrimination task and the objective probability of an event. A visual discrimination task was selected because it did not contain any explicit probability measures, because it provided a fairly critical test of the method, and because it permitted unambiguous control of the stimulus. Both experiments were conducted at Connecticut College[5] using the same 23 undergraduate women.

Each S was given a set of 45 problems from which her utility scale for points could be determined by linear programing (Davidson). A problem consisted of a choice between two alternatives. Each alternative consisted of winning 1500, 1000, 500, or 0 points or of losing 500 or 1000 points if Event ZEJ occurred; and winning or losing one of the above point scores if Event ZOJ occurred. The subjective probability of ZEJ (ZOJ) was assumed to be equal to 1/2. Linear programing solution could not

[5] The assistance of Mortimer H. Appelzweig and George O. Moeller in providing Connecticut College Laboratory facilities and students is gratefully acknowledged.

TABLE 1

DERIVED UTILITY SCALES

Subject	Objective Point Offers						ϕ
	−1000	−500	0[a]	500[a]	1000	1500	
QM	−1875	−1000	0	500	1375	1625	625
AG	−1061	−643	0	500	928	1142	71
MS	−2250	−1000	0	500	1000	1250	250
JN	−1500	−833	0	500	1000	1667	167
BW	−1333	−667	0	500	750	1083	83
GH	−1000	−750	0	500	750	1500	125
KT	−1333	−667	0	500	833	1333	167
CS	−1000	−500	0	500	875	1125	125
LB	−1250	−500	0	500	750	1250	250
EJ	−1300	−700	0	500	800	1200	100

[a] The zero points of all utility scales were arbitrarily set at zero objective points. The units of all utility scales were arbitrarily set by assigning 500 units to the distance between the utilities of zero and 500 objective points.

be found for many *S*s. The scales for 10 of the *S*s for whom solutions were found are shown in Table 1. Although the subjective distances between the point offers differed from the objective distances for most *S*s the rank-order of the points and the rank-order of the probabilities at which alternatives were preferred did not differ from the objective for the alternatives used in the two experiments.

Experiment I: Visual Discrimination

In the first experiment, *S* chose one of five alternatives on the basis of her ability to discriminate which of two line clusters contained the greater number of lines. Unknown to *S*, one of the two clusters always contained 100 lines. The other contained 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, or 125 lines. The lines were 1/8 to 1/4 inch long and were randomly distributed in an irregular, egg shaped pattern about 3/4 inch high and 1½ inches long. Both clusters were of the same irregular shape and area. Sixty pairs of these clusters were reproduced in mimeographed test booklets. Each booklet contained the same 10 pages with six problem pairs per page, but the order of the pages was randomized from one *S* to the next. The top cluster in each problem pair was labeled QUG and the bottom cluster was labeled QOB. Two of the QUG areas contained 75 lines, two contained 80, 85, 115, 120, and 125 lines. Four of the QUG areas contained 90, 95, 105, and 110 lines. Each of the 28 QUG areas was paired with a QOB area containing 100 lines. An additional 28 problems reversed the QOB and QUG areas. Four other problems were also presented in which QUG and QOB each contained 100 lines.

S was told that the test was designed to test her ability to estimate the number of objects in an area. The instructions informed *S* that she would be paid 1 cent for every 500 points she won, and that she would lose 1 cent for every 500 points she lost. However, *S* was guaranteed to net at least 1 dollar regardless of her losses. She could, if she selected properly, win up to 2 dollars. *S* won points by choosing one of the five alternatives described below for each pair of QUG and QOB areas.

Alternative CIJ: You win 1500 points if there are more lines in area QUG than there are in area QOB, and you lose 1000 points if there are more lines in area QOB.

Alternative MEQ: You win 1000 points if there are more lines in area QUG than in QOB, and you lose 0 points if there are more lines in QOB.

Alternative ZEJ: You win 500 points for sure.

Alternative XIP: You win 1000 points if there are more lines in area QOB than in QUG, and you lose 0 points if there are more lines in QUG.

Alternative FEH: You win 1500 points if there are more lines in area QOB than in QUG, and you lose 1000 points if there are more lines in QUG.

These alternatives are shown in Table 2, together with the alternatives offered in the next two experiments.

Two lines were drawn to the right of each problem in the test booklet. One line was labeled Side A, and the other was labeled Side B. *S* wrote the name of the alternative she preferred on the Side A line. If she wished, she wrote the same name on the Side B line, or she wrote the name of a different alternative. *S* was told that only one of the alternatives indicated on the two lines would be considered in computing her score. This alternative was determined by flipping a coin labeled Side A on one side, and Side B on the other. If *S* wrote the same name on both lines, she was sure to have that alternative count towards her score. If she wrote different names on the two lines, one or the other would be selected, each with a 50/50 chance. This procedure distinguished between real preferences for an alternative and those forced by a choice between equally desirable alternatives. This method provides a seven-interval probability scale as shown in Table 3.

In order to avoid changing *S*'s utility for points during the experiment, *S* was not told which area contained more lines, which side of the coin appeared face up, or her point

TABLE 2

ALTERNATIVES OFFERED THE SUBJECT BY PROBLEM AND CONDITION

Problems	Condition	Alternatives				
	S chooses column alternative..........	CIJ	MEQ	ZEJ	XIP	FEH
Discrimination	If top area contains more lines, *S* receives	1500	1000	500	0	−1000
	If bottom area contains more lines, *S* receives	−1000	0	500	1000	1500
	S chooses column alternative..........	VUK	WUB	XEH	YOF	
First set Probability Deck	If ZEJ card is drawn, *S* receives	1500	1000	500	−1000	
	If ZOJ card is drawn, *S* receives	−1000	500	1000	1500	
	S chooses column alternative..........	QJF	FHJ	ZBJ	DJQ	
Second Set Probability Deck	If XIH card is drawn, *S* receives	−1000	0	1000	1500	
	If XEH card is drawn, *S* receives	1500	1000	0	−1000	

TABLE 3

PROBABILITIES ASSOCIATED WITH DISCRIMINATION RESPONSES

Response		Probability QUG contains more lines										
			Subjective (utility scale)									
One side	Other side	Rational (Obj. Pts.)	QM	AG	MS	JN	BW	GH	KT	CS	LB	EJ
FEH	FEH	0–.33	0–.12	0–.19	0–.10	0–.31	0–.29	0–.43	0–.27	0–.20	0–.29	0–.24
FEH	XIP	.33	.12	.19	.10	.31	.29	.43	.27	.20	.29	.24
XIP	XIP	.33–.50	.12–.50	.19–.50	.10–.50	.31–.50	.29–.50	.43–.50	.27–.50	.20–.50	.29–.50	.24–.50
FEH or XIP or ZEJ	CIJ or MEQ or ZEJ	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
MEQ	MEQ	.50–.67	.50–.88	.50–.81	.50–.90	.50–.69	.50–.71	.50–.57	.50–.73	.50–.80	.50–.71	.50–.76
MEQ	CIJ	.67	.88	.81	.90	.69	.71	.57	.73	.80	.71	.76
CIJ	CIJ	.67–1.00	.88–1.00	.81–1.00	.90–1.00	.69–1.00	.71–1.00	.57–1.00	73–1.00	.80–1.00	.71–1.00	.76–1.00

score until she had completed all three phases of the experiments. The utility scale, the discrimination and the probability deck problems, discussed in the next section, were all administered during one group session that lasted 1½–2 hours.

If the SEU model and the method for measuring subjective probability are appropriate, we expect *S* to choose a high probability alternative (CIJ–CIJ) whenever she is sure that QUG contains more lines, to choose FEH–FEH whenever she is sure QUG contains less lines, and to vary her choice in the seven intermediate steps shown in Table 3 as her confidence in her ability to discriminate varies. One also expects a high correlation between *S*'s confidence, as measured by the subjective probability associated with her choice, and the difficulty of the discrimination task. Since the ranks of the subjective probability levels do not differ from the objective probability levels, for the 10 *S*s for whom utility scales were obtained (Table 3), a high correlation would be expected between the ranks of the objective probabilities associated with alternatives selected by *S* and the difficulty of the discrimination task. The magnitude of the difference between the number of lines in the two areas provides an objective basis for ranking the difficulty of the discrimination. Thus, if a significant correlation between the ranks of the stimulus differences and the ranks of the probability intervals associated with the alternatives is found, it would support the notion of using the proposed method to measure subjective probability, confidence in decisions, and stimulus ambiguity or difficulty.

Results. Two Kendall coefficients of concordance were calculated for each *S*, one for four replications of the five problems involving 100, 105, and

110 lines in QUG and 100, 105, and 110 lines in QOB; and the other for the four replications of the five problems involving 90, 95, and 100 lines in QUG and QOB.

Spearman average rank correlations obtained from these coefficients are shown, for each *S*, in Table 4. The "best" estimates of each *S*'s choice, on each problem, obtained by averaging the *S*'s four choices on each problem, were used to compute a coefficient of concordance for the group as a whole. The average Spearman rank correlation obtained from the coefficient is labeled Total in Table 4. Note that 29 of the 46 correlations are significant beyond the .05 level and that the totals are significant beyond the .01 level.

TABLE 4

CORRELATIONS BETWEEN OBJECTIVE STIMULUS SCALE AND SUBJECTIVE PROBABILITY

Subject	Task			
	Discrimination (5 alternatives)		Probability (4 alternatives)	
	One area contains more than 100 (100–105–110)	One area contains less than 100 (90–95–100)	First Set	Second Set
1	.38	.64**	.78**	.79**
2	.24	.71**	.74**	.78**
3	.30	.66**	.81**	.88**
4	.84**	.55*	.83**	.89**
5	.46*	.08	.83**	.40**
6	.70**	.61**	.87**	.87**
7	−.03	−.20	.68**	.55**
8	.24	.59**	.84**	.80**
9	.39	.57*	.86**	.30*
10	.64**	.30	.82**	.79**
11	.73**	.99**	.89**	.89**
12	.67**	.71**	.85**	.78**
13	.77**	.75**	.84**	.75**
14	.68**	.59**	.91**	.86**
15	.15	.47*	.89**	.89**
16	.11	.39	.56**	.66**
17	.61**	.18	.89**	.61**
18	.55*	.75**	.86**	.86**
19	.10	.39	.88**	.82**
20	.51*	.67**	.85**	.78**
21	.46*	.89**	.84**	.82**
22	.83**	.53*	.87**	.90**
23	−.10	.53*	.90**	.75**
Total	.834**	.871**	.87**	.76**

* Significant at .05 level.
** Significant at .01 level.

EXPERIMENT II: OBJECTIVE PROBABILITY

The results of a second pair of tests, which employed explicit probabilities rather than line areas, are also shown in Table 4. *S* was told that one card would be drawn from a deck containing 100 cards. Each card in the deck was labeled either ZEJ or ZOJ in the first set, and either XIH or XEH in the second set. *S* was told exactly how many ZEJ (or XIH) cards were in a particular deck. In the first set, she selected one of the four "Probability Deck" alternatives shown in Table 2 and described below:

VUK: You win 1500 points if ZEJ is drawn, and you lose 1000 points if ZOJ is drawn.

WUB: You win 1000 points if ZEJ is drawn and you win 500 points if ZOJ is drawn.

XEH: You win 500 points if ZEJ is drawn and you win 1000 points if ZOJ is drawn.

YOF: You lose 1000 points if ZEJ is drawn and you win 1500 points if ZOJ is drawn.

The objective probabilities at which these alternatives intersect are shown in Table 5. In the second set, she selected one of the four alternatives, shown in Table 2 and described below:

QJF: You lose 1000 points if XIH is drawn and you win 1500 points if XEH is drawn.

FHJ: You win 0 points if XIH is drawn and you win 1000 points if XEH is drawn.

ZBJ: You win 1000 points if XIH is drawn and you win 0 points if XEH is drawn.

DJQ: You win 1500 points if XIH is drawn and you lose 1000 points if XEH is drawn.

The probabilities at which these alternatives intersect are shown in Table 5.

TABLE 5

THE PROBABILITIES AT WHICH ALTERNATIVES INTERSECT

Response		Rational (Obj. Pts.)	Subjective (utility scale)									
One side	Other side		QM	AG	MS	JN	BW	GH	KT	CS	LB	EJ
						Probability ZEJ will be drawn						
YOF	YOF	0– .33	0– .10	0– .14	0– .08	0– .25	0– .20	0– .33	0– .21	0– .14	0– .22	0– .18
YOF	XEH	.33	.10	.14	.08	.25	.20	.33	.21	.14	.22	.18
XEH	XEH	.33– .50	.10– .50	.14– .50	.08– .50	.25– .50	.20– .50	.33– .50	.21– .50	.14– .50	.22– .50	.18– .50
YOF or XEH	VUK or WUB	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
WUB	WUB	.50– .67	.50– .90	.50– .86	.50– .92	.50– .75	.50– .80	.50– .67	.50– .79	.50– .86	.50– .78	.50– .82
WUB	VUK	.67	.90	.86	.92	.75	.80	.67	.79	.86	.78	.82
VUK	VUK	.67–1.00	.90–1.00	.86–1.00	.92–1.00	.75–1.00	.80–1.00	.67–1.00	.79–1.00	.86–1.00	.78–1.00	.82–1.00
						Probability XIH will be drawn						
QJF	QJF	0– .33	0– .12	0– .19	0– .10	0– .31	0– .29	0– .43	0– .27	0– .20	0– .29	0– .24
QJF	FHJ	.33	.12	.19	.10	.31	.29	.43	.27	.20	.29	.24
FHJ	FHJ	.33– .50	.12– .50	.19– .50	.10– .50	.31– .50	.29– .50	.43– .50	.27– .50	.20– .50	.29– .50	.24– .50
QJF or FHJ	ZBJ or DJQ	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50
ZBJ	ZBJ	.50– .67	.50– .88	.50– .81	.50– .90	.50– .69	.50– .71	.50– .57	.50– .73	.50– .80	.50– .71	.50– .76
ZBJ	DJQ	.67	.88	.81	.90	.69	.71	.57	.73	.80	.71	.76
DJQ	DJQ	.67–1.00	.88–1.00	.81–1.00	.90–1.00	.69–1.00	.71–1.00	.57–1.00	.73–1.00	.80–1.00	.71–1.00	.76–1.00

Each set of problems was constructed from 19 different decks. These decks contained the following numbers of ZEJ–ZOJ (XIH–XEH) cards: 0–100, 10–90, 15–85, 20–80, · · ·, 75–25, 80–20, 85–15, 90–10, 100–0. Thirty problems were constructed from each set of 19 decks by presenting 11 decks twice and 8 decks[6] once. Each set of 30 problems was included on a separate page in the test booklet after the discrimination experiment problems and the utility scale problems.

S again indicated the one, or two, alternatives she preferred, using the same Side A and Side B procedure. She was not informed of her point score on either set, nor of her point score on the discrimination or utility problems during the experiment.

Results. Kendall tau correlations between the objective probability ranks of the decks and the ranks of the alternatives selected by *S*, assuming *S* is rational, were computed for each *S* on each of the two problem sets. The results are shown in Table 4. The correlations labeled Total are average Spearman rank correlations obtained from Kendall's coefficient of concordance. The four entries in each row in Table 4 were obtained from the same *S* during the same experimental session.

Note that all 46 tau correlations computed for the probability problems are significant beyond the .05 level, and 45 of the 46 are significant beyond the .01 level. The average Spearman rank correlations are .87 for the first set and .76 for the second set. These results indicate that *S*s tended to select high subjective probability alternatives for highly likely events, and low subjective probability alternatives for events of low probability.

Summary of Experimental Results

The empirical results in both the discrimination and probability tasks support the use of the proposed method as a means for measuring subjective probability, confidence, and bias, and suggest that the method may provide a means for measuring stimulus ambiguity. The high correlations also indicate that it is an acceptable procedure for use in individual prediction. The method provides an estimate of subjective probability from a single response.

The fact that the correlations are less than 1.0, however, indicates that *S*s do not consistently choose the same response to the same stimulus. This may reflect fluctuations in utility during the experiment, or may result from obtaining subjective probability measurements that were more precise than those employed by *S*, or the results may be due to real changes in *S*'s expectancy from one stimulus presentation to the next. The fluctuations may further have been due, in this experiment, to the alternatives having approximately equal SEU over a wide range of discrimination problems and objective probabilities.

CONCLUSIONS

This paper described several methods for obtaining measures of subjective probability from a single response. One method required a priori knowledge of *S*'s utility scale and the other required that *S* never rank a set of alternatives inconsistently. Both methods were based on modifications of the SEU model and required that utility and subjective probability not interact.

Two experiments were conducted

---

[6] The eight decks presented only once were the 100/0, 85/15, 80/20, 70/30, 30/70, 15/85, 10/90, and 0/100.

to test one of the assumptions basic to both methods: the assumption that people vary their preference for alternatives according to subjective probability. The results supported this assumption, indicating that if utility and subjective probability do not interact, estimates of subjective probability can be obtained from a single response.

Several questions remain. The methods must be subjected to further empirical test. The reliability of the subjective probability estimates, the effects of increasing the number of alternatives presented to *S*, the relationship between subjective and objective probability, and the validity of the basic assumptions of the SEU model are still to be determined.

REFERENCES

Coombs, C. H. Inconsistency of preferences as a measure of psychological distance. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories.* New York: Wiley, 1959.

Coombs, C. H., & Komorita, S. S. Measuring utility of money through decision-making. *Amer. J. Psychol.*, 1958, **71**, 383–389.

Davidson, D., & Marschak, J. Experimental tests of a stochastic decision theory. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories.* New York: Wiley, 1959.

Davidson, D., Suppes, P., & Siegel, S. *Decision-making: An experimental approach.* Stanford: Stanford Univer. Press, 1957.

Edwards, W. The prediction of decision among bets. *J. exp. Psychol.*, 1955, **51**, 201–214.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

Luce, R. D., & Raiffa, H. *Games and decisions: An introduction and critical survey.* New York: Wiley, 1957.

Savage, L. J. *The foundations of statistics.* New York: Wiley, 1954.

Siegel, S. A method for obtaining an ordered metric scale. *Psychometrika*, 1956, **21**, 207–216.

Suppes, P., & Walsh, K. A nonlinear model for the experimental measurement of utility. *Behav. Sci.*, 1959, **4**, 204–211.

(Received August 23, 1960)

Psychological Review
1962, Vol. 69, No. 2, 149–155

THEORETICAL NOTES

A NOTE ON INTENSITY GENERALIZATION AND PROTHETIC SCALING[1]

SHELDON H. WHITE

The University of Chicago

Tests for stimulus generalization to variation in stimulus intensity have generally followed a typical procedure. A set of three, four, or five stimulus intensity values are selected. One group of subjects learns a response to the weakest intensity, another group learns to the strongest, and then both groups are tested for generalized responding to the complete set of stimuli.

The two-way gradients found by such a procedure have shown a typical asymmetry despite a respectable diversity of types of subject, stimulus dimensions, and learning paradigms. A description of the asymmetry given by Hull (1949) notes five points: (*a*) The curve of the group trained at the weak intensity value is concave downward; (*b*) The curve of the strong-trained group is concave upward; (*c*) The strong-trained group shows stronger response to its training stimulus than does the weak-trained group; (*d*) The downward curve is steeper in slope than the upward curve; (*e*) The upward and downward curves intersect at their midpoints.

Explanations of the asymmetry have involved an energizing property of CS intensity (Hull, 1949), or else a differential conditioning process involving contextual stimuli (Logan, 1954; Perkins, 1953). However, the shape which a curve will appear to have depends, of course, upon baseline and ordinate scaling. The purpose of the present note is to point out that: (*a*) newer psychophysical scales offer an alternative to the practice, traditional in intensity generalization studies, of selecting intensity values at equal logarithmic intervals; (*b*) arbitrary logarithmic scaling can produce generalization asymmetry resembling the typical picture under discussion; and (*c*) rescaling baselines of existing curves in terms of magnitude-estimation scales reduces the apparent asymmetry of some of the intensity generalization data.

SCALING OF PROTHETIC DIMENSIONS

Stevens (Stevens, 1957; Stevens & Galanter, 1957) has recently presented evidence for a dichotomy among stimulus dimensions. Those dimensions along which stimulus change appears to take place by addition or subtraction of stimulation are classed as prothetic, while other dimensions, for which change is to some extent a matter of substitution, are classed as metathetic. Prothetic dimensions are generally associated with intensity attributes. The stimulus dimensions along which asymmetry of generalization has been found (brightness, loudness, and area) are classified as prothetic, while those which have showed approximate symmetry (hue, pitch, locus on skin, locus in space, length of line) are either metathetic *or have been treated as though they were*. That is, experimental sets of stimulus values have not been logarithmically spaced in the manner usually employed for prothetic dimensions.

The logarithmic spacing of a series of intensity values arranges them so that approximately equal numbers of jnd's fall between successive values in the series, assuming that the stimuli so spaced are thereby also arranged in even steps of similarity, or psychological distance. The plausibility of the assumption has led investigators of stimulus generalization to space their stimuli in accord with the principle; Hovland (1937a) actually counted jnd's to locate stimuli, and later studies have used the logarithmic spacing as a good approximation of this procedure.

---

[1] Study performed with the aid of a Ford Foundation in Behavioral Sciences research grant. The writer wishes to express his gratitude to Donald W. Fiske, Austen H. Riesen, and Charles C. Spiker for helpful critical readings, and also to thank Paul Grim for computing the asymmetry indices of the two figures.

Another psychophysical estimate of apparent stimulus distances produces an alternative scaling. The method of magnitude estimation requires the observer to assign numbers to stimuli in accord with what he judges their intensities to be. Differences between numerical magnitude estimations for stimulus intensities are typically in better agreement with the actual physical energy differences between those stimuli than the jnd curve might lead one to expect. That is, magnitude estimations are generally related to physical energy values by a power function, of less curvature than the logarithmic jnd function. Some dimensions even show a linear relationship between magnitude estimation and physical value (Stevens & Galanter, 1957).

While logarithmic (jnd) scaling has been traditional and accepted, magnitude scaling has some points in its favor. The summing of jnd's to estimate psychological distance rests on an unproven assumption, historically controversial, that jnd's are equal, summable, "units" of sensation. Magnitude scaling does not appear to require such a critical assumption. Further, magnitude scaling, and only this psychophysical technique to date, has showed some amount of validity by permitting quantitative cross-modality scaling predictions (Stevens, Mack, & Stevens, 1960).

EFFECTS OF LOGARITHMIC SCALING

What is the effect of baseline scaling on a two-way generalization curve? Suppose that five intensity values have been chosen so that the distances between their logarithms are equal. The line at the top of Figure 1 illustrates the placement of the stimuli, A-E, on the physical

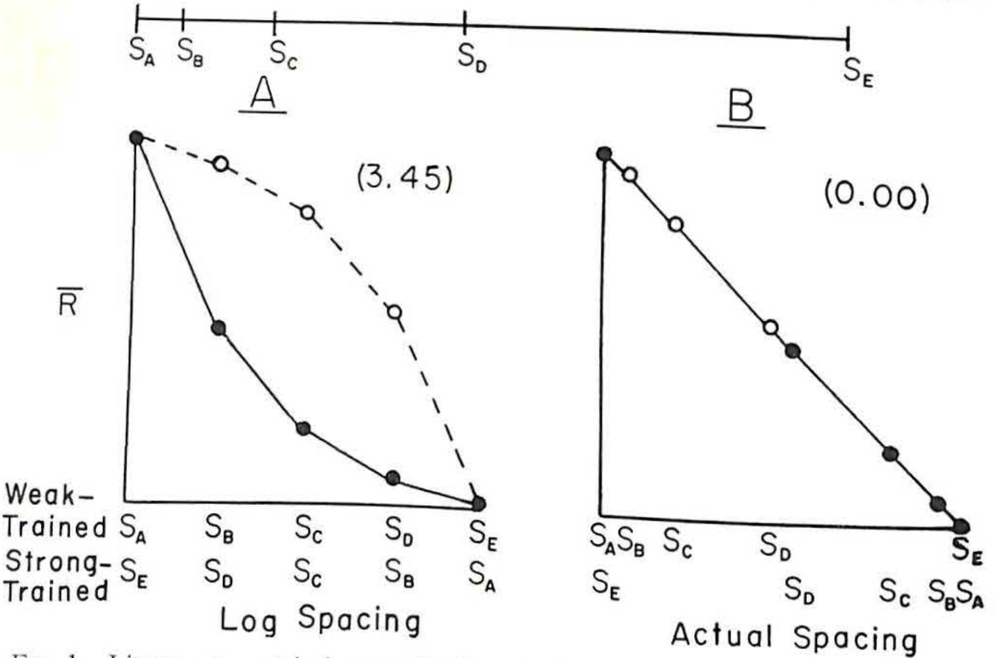


FIG. 1. Linear, symmetrical generalization curves inappropriately drawn over a logarithmically scaled baseline (Graph A), and drawn over the correctly scaled baseline (Graph B). (The training points for the weak-trained group—open circles, dashed line, and the strong-trained group—filled circles, solid line, are at the left, and the curves extend rightward. The parenthesized numbers are a rough index of asymmetry, representing the number of square inches between the curves as originally drawn in 3″ × 3″ size.)

dimension; the traditional assumption would be that they mark off approximately equal apparent steps.

A weak-trained group of subjects learns a response to S_A, while a strong-trained group learns to S_E, following which both groups are tested for generalized responding to the complete set of five stimuli. We will suppose that the generalization function is linearly, symmetrically related to physical energy differences between training and test stimulus; the subject generalizes in direct proportion to the physical energy differences between stimuli.

Hypothetical results under the above conditions are given in Figure 1, in both the A and B graphs. In the A graph, the spacing of the stimuli along the abscissa is even, following the assumption that equal logarithmic distances represent equal apparent stimulus change. The generalization curves appear to be markedly asymmetrical, with that for the weak-trained group being concave downward and that for the strong-trained group being concave upward.

In the B graph, the stimuli are spaced according to their physical energy distances, and the curves for the weak-trained and strong-trained groups are symmetrical and linear.

It is possible that the practice of logarithmic spacing, characteristic of intensity generalization studies, may have analogously led to an observed asymmetry. Some support for this argument will now be sought in the existing literature.

The analysis of the hypothetical example just given is to serve as a model for what follows. In that example, it was assumed that apparent stimulus differences were in exact accord with the physical energy differences between stimuli. For the studies to be considered, it will be assumed that apparent stimulus differences are best estimated from the scales yielded by the psychophysical method of magnitude estimation.

STUDIES OF INTENSITY GENERALIZATION

The general format of a study of bidirectional intensity generalization has been described above. Some distinction should be made between a "pure" design—in which each subject is tested on only one stimulus for generalization—and the more common counterbalanced design, where each subject is tested on all of the stimuli, with different subgroups of subjects used to counterbalance the order of testing among stimuli. Despite the fact that the data from counterbalanced designs are more difficult to interpret—being a mixture of generalization, extinction, retraining, and generalization of extinction effects—such designs have been used because of the considerable economy in number of subjects which they permit.

The relevant literature may now be briefly described:

1. Hovland (1937a) used a GSR conditioning procedure, with his stimuli being four 1000 cycle tones separated in loudness by 50 jnd steps. Separate groups of 16 college students were trained at his faintest and loudest tones, and then tested by a counterbalanced procedure.
2. Brown (1942) used rats in a straight alley approach situation. His stimuli were three illuminations of a large ground-glass screen at the end of the alley. There were two groups of 18 rats each, tested by a counterbalanced procedure.
3. Grice and Saltz (1950) also used rats in a straight alley approach situation. Their stimuli were five white circles, four of them graded in area at logarithmically equal intervals and the fifth at an intervening half-step. Testing was by a pure procedure, with 15 subjects for each test stimulus. Two sets of data are to be considered here, the resistance to extinction data reported originally, and speed data reported later (Grice, 1956).
4. Miller and Greene (1954) used rats in a straight runway avoidance paradigm. The stimuli were three loudnesses of a buzzer. There were 30 rats, tested by a pure procedure. Each data point of the resistance to extinction measure was based upon five animals.
5. Spiker (1956) tested children on a task in which they responded freely dur-

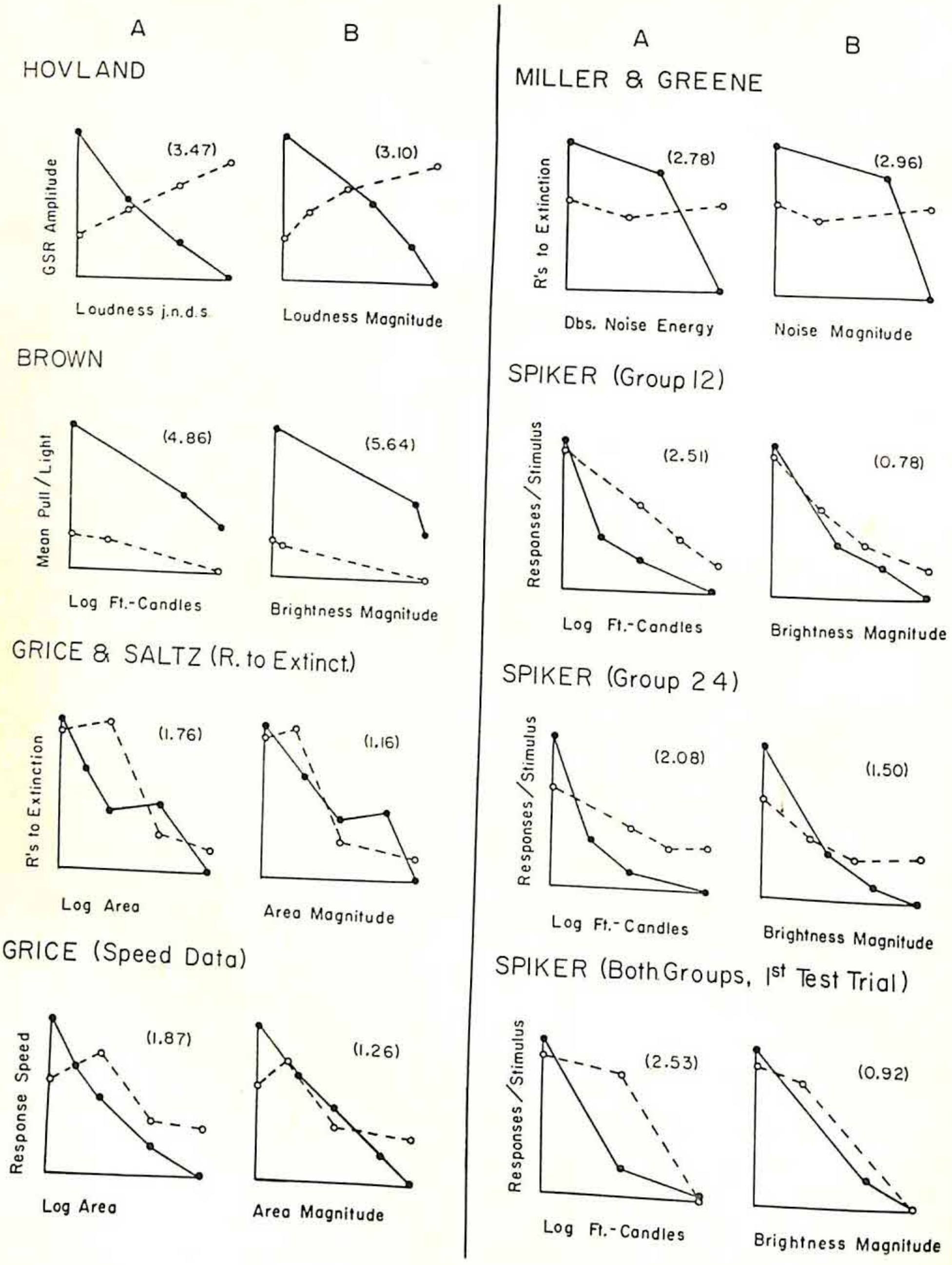


FIG. 2. The A columns show the results of intensity generalization studies with abscissae scaled to log stimulus energy (or, in Hovland's case, jnd distances). The training points for the weak-trained groups (open circles, dashed lines), and the strong-trained groups (filled circles, solid lines) are at the left, and all curves extend rightward. In the B columns, the same data are presented over abscissae rescaled in terms of psychophysical magnitude estimations. (All figures have been proportioned to occupy the same square area. The parenthesized numbers are a rough index of asymmetry, representing the number of square inches between the curves as originally drawn in 3″ × 3″ size.)

ing 3-second trial intervals. His stimuli were four illuminations of a circular ground-glass stimulus window; three of the stimuli were at logarithmically equal foot-candle intervals and the fourth at an intervening half-step. Thirty children were given 12 trials of training (15 weak-trained and 15 strong-trained), followed by a counterbalanced testing procedure; another 30 were given 24 trials of training before test. These two training groups will be considered separately below. In addition, the data of the pooled groups on their first generalization test trial will be considered, as a relatively pure generalization measure.

An additional study (Heyman, 1957) is clearly relevant in design but will not be further considered because the data, as tabled, did not permit clear appraisal of curve shapes.

The A columns of Figure 2 show the generalization curves reported for the above studies. To provide a clearer picture of the degree of symmetry or asymmetry of the curves, the training points of both the weak-trained and strong-trained groups have been placed on the left, and the curves extended rightward. All figures have been proportioned to occupy the same square area. The scaling along the abscissae of the Column A figures is logarithmic, except in the case of Hovland, where it is in terms of jnd units.

The ordinates of the Column B figures are identical with those of their Column A counterparts; the abscissae of the Column B figures, however, have been rescaled in the terms of psychophysical magnitude estimations. The stimulus energy values (E) for each study were transformed to magnitude scale values (M) using the power function $M = E^k$, where the exponent k is a characteristic of the dimension established by psychophysical investigation. For Hovland's loudnesses of 1000 cycle tones, k was 0.3 (Stevens, 1956). For Brown's study, as for Spiker's, brightness of illumination called for a k of 0.36 (Stevens & Galanter, 1957). Miller and Greene's buzzer loudnesses were converted using an equation for white noise (Stevens, 1958) with a k of 0.3. Finally, Grice's dimension of visual area involved a k range of 0.9–1.15 (Stevens, 1957); a k of 1.0 was used here.

If the asymmetry of the Column A figures were the result of baseline scaling alone, their treatment in Column B should make them roughly symmetrical. Several extraneous factors in the data being considered combine to make this a fairly gross specification.

First, the psychophysical data on which the rescalings were based comes from adult human observers, but three of the five studies were done with rats and a fourth with children. The Column B baselines are therefore only educated guesses at a proper scale for those subjects.

Second, the response strengths of the weak-trained and strong-trained groups were not always equal at their respective training stimuli. Generalization curves extending from training stimuli of different height would not be superimposed, nor is there any good reason to expect them to be parallel. Connected in here is an ordinate scaling problem—the probability that the same amount of response change has different implications starting from different levels of responding.

A third factor which might prevent rescaling from bringing about full symmetry applies particularly to counterbalanced designs. If the line at the top of Figure 1 is examined, it will be seen that logarithmic spacing of stimuli causes those of weaker intensity to be placed relatively closer together than stimuli of greater intensity. During repeated testing, the net effect of this is to produce less generalization of inhibition at the training value for the strong-trained group than for the weak-trained group, while loading relatively more generalized inhibition onto the tail of the strong-trained group. The strong-trained group consequently shows a steeper gradient than does the weak-trained group. This difference in steepness is not removed by rescaling. It might help account for differences between bright-trained and dim-trained subjects at their respective training values, the factor discussed in the preceding paragraph.

Now, examining Figure 2B, it is seen that the rescaling meets with partial success in symmetrizing the curves. Five of the curves—the two derived from the Grice data, and the three derived from the Spiker study—appear to be brought closer to symmetry. These five curves represent only three completely independent sets of data among them. On the other hand, the rescaling of the Hovland, Brown, and Miller and Greene studies has worked no improvement.

DISCUSSION OF STUDIES

Despite the fact that Hull advanced his five rubrics from knowledge of only the Brown and Hovland data, the failure to symmetrize either these or the Miller and Greene data is, to some degree, tied to discrepancies between these studies and Hull's typical picture. Hovland's weak-trained group shows a generalization curve regularly ascending away from the training point; Brown's weak-trained and strong-trained curves show no overlap; the Miller curves are exactly opposite in curvature to the standard. Only in the case of the Miller and Greene data—for which N is relatively low and reported response variance quite high—is there a clear cause to doubt the reliability of the reported curve shapes. Indeed, the ascending trend of Hovland's weak-trained group was reproduced by him in two later sets of intensity GSR generalization data (Hovland, 1937b).

On the other hand, the Grice and Saltz, and Spiker data, which on the whole provide the best complete approximation to Hull's picture, are most altered by rescaling. Their residual asymmetry might be due to the extraneous factors noted above, to sampling error, or to an intensity effect which is only reduced in apparent magnitude by rescaling.

Differentiae of studies not made symmetrical by rescaling might include: (*a*) a large difference between response strengths of the dim-trained and bright-trained groups to their training stimuli; (*b*) aversive, as opposed to appetitional, incentives; (*c*) auditory, as opposed to visual, stimuli; (*d*) counterbalanced as opposed to pure, designs; or (*e*) a stimulus set of greater "impact," whether because of intensity or of dominance as figure against ground.

What can be said, then, about the rescaling venture? There is no a priori reason for preferring one expression or picturing of a set of data to another, but choices can be made in terms of usefulness for interpretation and communication. Here, the limited reduction of asymmetry by magnitude scaling may usefully clarify where and how much a special intensity factor must be invoked.

There is still evidence for an asymmetry of generalization, which may reflect an intensity factor. However, the intensity interpretation still faces as a difficulty the fact that a majority of the half-dozen conditioning studies which have searched for an effect of CS intensity have been unable to find it.

SUMMARY

The typical asymmetry found in studies of two-way intensity generalization may involve baseline scaling, since all such studies have spaced their stimuli at equal jnd distances, either directly or by a logarithmic approximation. Magnitude estimations of stimuli yield a different but justifiable power-function spacing. An artificial example of a typical asymmetry produced by logarithmic spacing is given. Rescaling of intensity generalization data in the literature to magnitude-scale baselines results in increased symmetry in some cases.

REFERENCES

BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.

GRICE, G. R. Response speed and size generalization. *Psychol. Rep.*, 1956, **2**, 246.

GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–709.

HEYMAN, W. Certain relationships between stimulus intensity and stimulus generalization. *J. exp. Psychol.*, 1957, **53**, 239–248.

HOVLAND, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291. (a)

HOVLAND, C. I. The generalization of conditioned responses: III. Extinction, spontaneous recovery, and disinhibition of conditioned and of generalized responses. *J. exp. Psychol.*, 1937, **21**, 47–62. (b)

HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

LOGAN, F. A. A note on stimulus intensity dynamism (*V*). *Psychol. Rev.*, 1954, **61**, 77–80.

MILLER, W. C., & GREENE, J. E. Generalization of an avoidance response to varying intensities of sound. *J. comp. physiol. Psychol.*, 1954, **47**, 136–139.

PERKINS, C. C. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225–231.

SPIKER, C. C. The stimulus generalization gradient as a function of the intensity of stimulus lights. *Child Develpm.*, 1956, **27**, 85–98.

STEVENS, J. C. Stimulus spacing and the judgment of loudness. *J. exp. Psychol.*, 1958, **56**, 246–250.

STEVENS, J. C., MACK, J. D., & STEVENS, S. S. Growth of sensation on seven continua as measured by force of handgrip. *J. exp. Psychol.*, 1960, **59**, 60–67.

STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

(Received December 19, 1960)

Psychological Review
1962, Vol. 69, No. 2, 156-160

ALL-OR-NONE VERSUS INCREMENTAL LEARNING[1]

JOAN E. JONES

Defence Research Medical Laboratories; Toronto, Canada

The process of learning may be conceived of as the gradual strengthening of connections between events or, alternatively, as the establishment of such connections in an all-or-none fashion. Estes (1960) presented results which suggest that evidence for the all-or-none position tends to be obscured by the traditional techniques of grouping data. In one of a number of miniature experiments cited by Estes (1960, pp. 213–215) paired associates were recalled on successive test trials (T_1, T_2) after a single reinforced presentation (R_1). The sequence was R_1, T_1, rest, T_2. Estes argued that, if an incremental theory is tenable, the probability of a correct response appearing on T_2 should be independent of recall on T_1. An all-or-none theory suggests that correct responses on T_2 should occur only by chance unless a correct response was given on T_1 (see Figure 1). A close relationship was found between recall on T_1 and T_2, i.e., of the 49% of cases correct on T_1 75% were correct on T_2 (C-C cases) and of the 51% of cases incorrect on T_1 91% were incorrect on T_2 (N-N cases).

Estes' statement of incremental theory forces upon it the assumptions that a reinforced presentation results in equal increments in associative strength for all items in all subjects and that the strength of competing responses is equal in all cases. His deductions concerning the probability of a correct response on T_2 do not necessarily follow if these assumptions are not made. Estes contends that he has disposed of the point of item difficulty and individual differences by demonstrating that for his subjects and with his items a second reinforcement (R_2) interpolated between T_1 and T_2 yields approximately the same percentage of newly correct responses on T_2 (46%) as occurred on T_1 following R_1 (40%) (Estes, 1960, pp. 215–216). However, any prediction of the relative percentages of newly correct responses following one reinforced presentation and following two, involves making assumptions about the distributions of item difficulty and learning ability. The character of those distributions for such a limited number of items and subjects can be known only empirically. Although the approximate equivalence of the obtained percentages is consistent with an all-or-none interpretation, the demonstration does not at all preclude the possibility that R_1 results in unequal increments for different cases.

According to the latter view, the associations have achieved unequal strengths after R_1, hence they do not have equal probabilities of appearance. Subclassifying the cases into those which do (C) and those which do not (N) appear on T_1 biases the probabilities of appearance on T_2 without further reinforcement in favor of the C cases. Assuming that there is no forgetting between T_1 and T_2, the C-N and N-C cases represent approximately half of those in which the associations are at about threshold strength and the number of cases in each should be approximately equal. (The proportions will be equal in the special case where the C/N split on T_1 is 50/50.) Forgetting is reflected in an increase in the number of C-N cases and a decrease in the number of N-C cases. That is, incremental theory does not have to predict the frequency of N-C cases. However, it can predict trends in the N-C and C-N categories when the conditions of the experiment produce greater or lesser degrees of forgetting. According to the all-or-none theory, forgetting will be reflected in an increase of cases in the C-N category but the frequency of cases in the N-C category will be unaffected unless subjects are guessing intelligently, in which case it will decrease.

---

[1] Defence Research Medical Laboratories Project No. 241, Report No. 241–4, PCC No. D77-94-01-16, H.R. No. 204.

Another problem which precludes uncritical acceptance of Estes' interpretation is the indeterminate theoretical chance level. According to an all-or-none theory the number of N-C cases should be no greater than that expected on the basis of sheer guessing. The chance value depends on the population of responses among which the subject is guessing, which might be a number less than, equal to or greater than eight, the number of items presented, and cannot be specified other than empirically. The alternative forms of guessing might be referred to as "intelligent," "unintelligent," and "stupid" guessing. Further information is necessary in order to evaluate whether the N-C cases which do occur, whatever their number, are sheer guesses or possible true recalls. This information is not available in the data as collected by Estes but an extension of his procedure may clarify why N-C cases do occur and also indicate whether associations have been established by R_1, though not revealed on T_1.

Three experiments are reported here. The first is a replication of Estes' experiment with the trial sequence R_1, T_1, "rest," T_2, confirming his result. The second is a variation and extension of Estes' procedure, the trial sequence being R_1, T_1, T_2, T_3, T_4. The third is a repetition of Experiment II using a double-blind technique.

EXPERIMENT I

Twenty-seven housewives acted as subjects in Experiment I. Each of eight associate-pairs, a consonant as stimulus and a number as response, were shown on separate 3" × 2½" cards. The responses were the numbers 1 through 8. The experimenter sat at a table opposite the subject and presented the cards manually after shuffling. On R_1 the cards were each held while the subject read aloud the consonant and the number. On test trials cards bearing only the consonant stimulus were shuffled and then presented. The test cards were held until the subject responded with the number she though correct or with a guess. No omissions were allowed. T_2 followed T_1 after approximately 7 minutes during which time the subject took part in another experiment which involved sorting cards bearing lines of different lengths and orientations. There was no warning that a second test would be made.

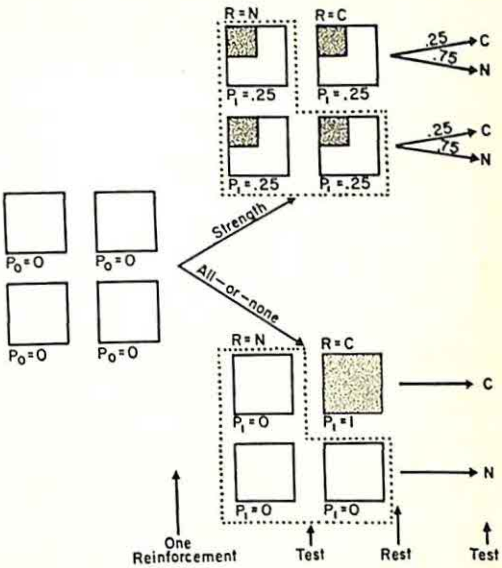


FIG. 1. Schema representing effects of a single reinforcement according to incremental (upper branch) vs. all-or-none (lower branch) theories. (Squares represent subjects with the proportion of darkened area in each indicating the probability of the correct response, C, for the given individual) (from Estes, 1960, p. 214.)

Experiment I was not an exact replication of Estes'. Apart from the usual differences of experimenter, subject, time, and place the present experiment differed from Estes' in four respects: (*a*) In the present experiment consonants rather than three-consonant syllables were used as stimuli. (*b*) Because the cards were shown manually the presentation and between-presentation duration were variable whereas

Estes timed both of these intervals. (*c*) No omissions were allowed whereas Estes was forced to accept some omissions because of the restricted intertrial interval. (*d*) The "rest" period between T_1 and T_2 was longer than that allowed by Estes and occupied with a different activity. From Estes' full report of his experiment (Estes, Hopkins, & Crothers, 1960) it appears that the "rest" interval was brief (estimated at 1½ minutes) and was used for a reinforced presentation of other items from the list. The probable unimportance of these procedural differences can be assessed from the results of Experiment I which are essentially the same as those obtained by Estes. Of the 53% of cases correct on T_1 75% were correct on T_2 (C-C) and of the 47% not correct on T_1 81% were not correct on T_2 (N-N).

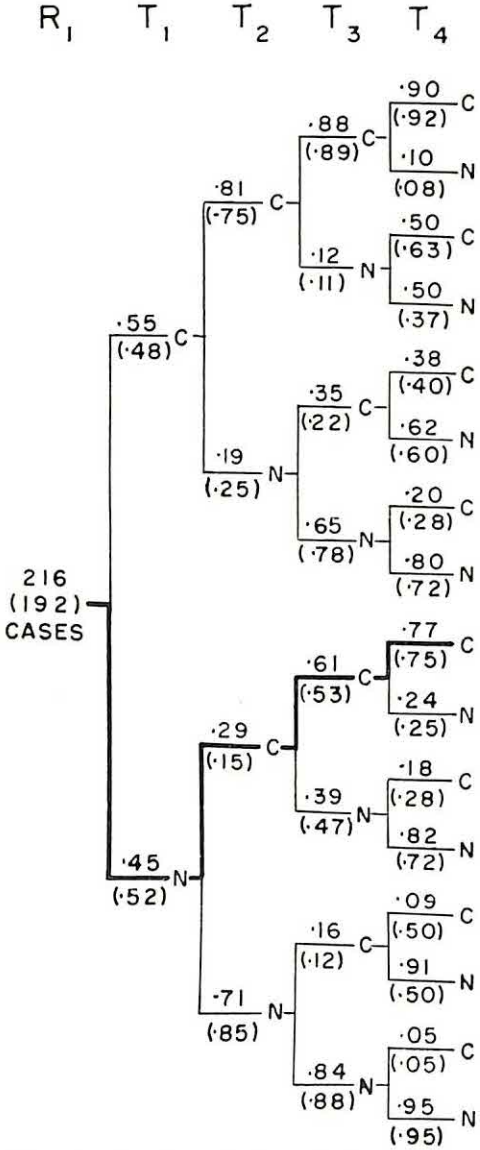


FIG. 2. Results of Experiment II and Experiment III (in parentheses). (Empirical values are proportions of instances in which correct, C, and incorrect, N, responses on a given test trial were followed by C and N responses on the following test trial.)

EXPERIMENT II

The subjects for Experiment II were the same 27 women who served in Experiment I. Both experiments were conducted in the same session, Experiment II first, and were separated by approximately 7 minutes during which time the subject took part in the card-sorting experiment. The procedure was as described for Experiment I except that the test cards were reshuffled and represented immediately until four test trials were completed. The experimenter gave no indication beforehand that more than one test would be made. The stimuli were consonants, none of which appeared in Experiment I. The responses were again the numbers 1 through 8.

The probability of recall on T_{2-4} largely depends on whether the correct response is given on T_1 (see Figure 2), again confirming Estes' findings. However, features of the results are difficult to reconcile with an all-or-none theory.

Although the percentage of N-C cases (29%) is well within the range of chance variation, it is necessary to explain the tendency to repeat those guesses on T_3 and T_4, i.e., N-C-C and N-C-C-C cases. One would expect all-or-none theory to require the N-C-C and N-C-C-C values to approximate .29, i.e., the guessing rate established on T_2. The incremental interpretation demands that those values approximate .50, these cases being at threshold strength. The obtained values of .61

TABLE 1

TRENDS IN C-N AND N-C CASES AS A FUNCTION OF FORGETTING AND A COMPARISON OF N-C VALUES WITH THOSE EXPECTED ON THE BASIS OF INTELLIGENT GUESSING (g)

Experiment	Percentage correct		Retention index T_2/T_1	C-N	N-C	g
	T_1	T_2				
III	.55	.57	1.04	.19	.29	.22
II	.53	.48	.91	.25	.19	.21
Estes	.49	.39	.80	.29	.09	.20

and .77 are too high for either of these interpretations. These cases cannot be fully accounted for as part of a general tendency to repeat earlier calls on later trials. Of those who change their call on T_2, having been wrong on T_1, 61% of those correct repeat the second call once and 36% twice while of those again incorrect only 37% repeat the second call once and 14% twice. If the N-C cases are regarded as guesses it is necessary to explain why those calls are repeated with a much higher probability than incorrect responses. If they are regarded as true recalls the expected N-C-C and N-C-C-C values are .50, or less (if forgetting occurs), but in fact the values increase.

The implication is that there is some kind of feedback which distinguishes past responses as correct or incorrect. One possible source of such information under the conditions of Experiment II is the experimenter, who knows whether the response is correct or not. Experiment III was conducted to check this possibility. A new group of 24 subjects drawn from the same population learned the items used in Experiment II. One experimenter presented the paired-associates on R_1 and a second experimenter, who was unaware of the correct associations, presented the test cards on T_1 to T_4. The results of Experiment III appear in brackets in Figure 2. Again a high proportion of calls correct for the first time on T_2 were repeated on T_3 and T_4—a proportion considerably in excess of repetitions of incorrect calls. Of those correct on T_2 53% repeat the response once and 40% twice, while of those incorrect only 32% repeat the call once and 12% twice.

If the feedback does not originate from an unwitting experimenter then it must be somehow induced by the response itself. Bearing in mind that the associations in question are assumed to be at threshold strength and also the fact that recognition is easier than recall, it seems not unreasonable to suggest that in some cases the subject may recognize that the response she has just given as a guess was actually correct. Such an interpretation renders quite conceivable the notion that learning can occur on test trials. Estes et al. (1960) in fact noted such a phenomenon. However, the interpretation suggested does require that an association of liminal strength be set up on R_1 and is therefore inconsistent with an all-or-none interpretation.

Further, comparison of the results of Estes' experiment with those of Experiments II and III shows an effect which is also handled more readily by an incremental theory. If the total number of correct responses is taken as an index of retention then it is apparent that forgetting between T_1 and T_2 is greatest in Estes' experiment and least in Experiment III (see Table 1). The trend is consistent with the fact that Estes' subjects learned other associations between trials while the subjects in Experiment II sorted cards and in Experiment III there was no interpolated activity. From Table 1 it is also seen that as forgetting increases so the proportion of cases in the C-N category increases, as predicted by both theoretical approaches, but also that the proportion in the N-C category decreases. A χ^2 test based on the frequencies in the N-C categories showed the latter effect to be significant at the .001 level of confidence. The effect cannot be accounted for in terms of an increase in the population of responses among which the subject is guessing. The expected N-C values for intelligent guessing were calculated as follows. If C is the proportion correct on T_1, c the probability that an item will be learned on R_1, and g the probability that the correct response will be guessed, then the equation for C (Estes et al., 1960) is:

$$C = c + (1 - c)g \qquad [1]$$

Substituting the obtained value of C and an arbitrary value of g (with the limitation that $g < C$) into Equation 1, solve for c. A new estimate of g is obtained from:

$$g = \frac{1}{n(1 - c)}$$

where n is the number of items in the list. The new estimate of g is then substituted into Equation 1 and the procedure reiterated until a stable value of g is obtained. The g values obtained after three iterations are shown in Table 1. A χ^2 test of the obtained N-C values against the frequencies predicted for intelligent guessing yielded a p of .001. That is, the drop in the N-C frequencies as a function of forgetting exceeds that expected simply as a decline in the chance level.

The data of Experiment II strongly suggest that subclassifying the cases according to their history on successive trials does no more than sort them in order of difficulty, each case being one item for one subject. The probability of appearance of a correct response on T_4 is: for cases formerly correct on all three trials (C-C-C cases), .90; for cases formerly correct on two of three occasions (C-C-N, C-N-C, and N-C-C), .50, .38, and .77 respectively; correct on one of three occasions (C-N-N-, N-C-N, and N-N-C), .20, .18, and .09 respectively; for those never before correct (N-N-N), .05. A similar ordering can be observed in the data of Experiment III. According to the all-or-none position the T_4 probabilities should be the same for all categories with the exception of C-C-C.

The writer appreciates Estes' statement that he does not mean to offer the experiment he reports as crucial tests of incremental versus all-or-none theory and agrees with his thesis that theorists should examine closely the extent to which their concepts are artifacts of experimental procedures. But insofar as the minature experiments are used as demonstrations, they are taken to create difficulties for incremental theory. Considered in the context of additional information from the experiments reported in this paper these same data raise difficulties for all-or-none theory. The criticism directed at the miniature experiments is not that they are too simple but that they are incomplete.

REFERENCES

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

(Received February 2, 1961)

VOL. 69, NO. 3 MAY 1962

PSYCHOLOGICAL REVIEW

ON THE STRUCTURE OF ASSOCIATIVE MEANING[1]

JAMES DEESE

Johns Hopkins University

Ever since Hobbes wrote "From St. Andrew the mind runneth to St. Peter and from St. Peter to *Stone*," psychologists and philosophers concerned with the nature of the mind have been trying to describe the nature of the associations that are so evident in human verbal behavior. Locke tried to picture the structure of the human mind by describing relations between associations. He attributed associations to the more or less accidental contingencies of perceptual qualities in the real world. By so doing, Locke became responsible for a major tradition in the theory of both thought and perception.

Almost all of the theoretical attempts to deal with association stem from Locke. For experimental psychologists interested in learning, however, the most important theoretical contributions come from the so-called secondary laws of association, introduced by Thomas Brown and David Hartley. Associations themselves are supposed to arise by contiguity, similarity, etc., but they occur in the strengths and distributions that they do because of their frequency, vividness, and so on. It is experimental treatment of the secondary principles that has led to most of the stable empirical generalizations about associative learning, though the primary "laws" have always been of greatest theoretical importance.

Since Ebbinghaus, the major experimental treatment of association has been with artificial material, mostly nonsense syllables. Consequently the relations of contiguity, similarity, etc., have been arranged by experimenters, not obtained from the natural relations among linguistic elements.

Somewhat earlier than Ebbinghaus, however, Sir Francis Galton inaugurated the study of naturally occurring associations. For the most part, the tradition begun by Galton has remained psychometric rather than experimental. Recently, these two traditions have been combined. Nevertheless, most of the attempts to deal with the structure or organization of naturally occurring associations have been empirical rather than theoretical, in the tradition of Locke.

Most of the empirical studies of the organization of naturally occurring associations have been studies of classification of associations. Associations are classified by logical relations or by the dictionary meanings of words which occur as stimuli or as responses in free association tests.

A good summary of most of the attempts to discover the structure of associations by classification can be found in the first edition of Woodworth's *Experimental Psychology* (1938). Woodworth both reviews a half century of German, American, and English attempts at classification and presents a classification scheme of his own. Woodworth's own classification is more obviously rooted in the association frequencies themselves than are the earlier ones. It is also unique in that it is a two dimensional classification. Woodworth classifies both by meaning, in the traditional sense, and by meaningfulness (in the sense used by Noble and others).

---

[1] The work reported in this paper was supported by funds from the National Science Foundation, Grant 13055.

As with the earlier classifications, Woodworth's is a combination of dictionary and grammar book rules, filtered through Woodworth's own verbal behavior and the verbal behavior of the investigators whose work he summarizes. The trouble with it and with all classifications to date is that they are not completely rooted in the association process itself. They attempt to impose upon the associations the logical relations found in grammars and the relations between words stated in dictionaries and thesaureses as well as some of the relations among objects perceived in the natural world. Perhaps the major result of such classifications is to convince us that while associations are partly related to the *meanings* of words in all of the above senses, associations are meaningful in precisely none of the above senses.

The meaningfulness of words refers to organized relations among the words and among words and objects in the natural world. Associations are related to one another in some organized fashion and also related in some way to perceived properties in the natural world. The only conclusion is that there is some sense in which one may speak of "associative meaning" and that the way to discover the associative meaning of words is not to classify associations according to other senses of meaning but to discover the relationships associations have with one another. That is the purpose of the present paper.

Aside from the difficulty summarized above, all previous attempts at classification, including the interesting one proposed by Karwoski and Berthold (1945), have suffered from another defect. All schemes proposed thus far have attempted to deal simultaneously with the stimuli and the responses to them. That is to say, they attempt to classify the *relation* between stimuli and responses, even though the individual members vary. Thus, the result frequently is an attempt to equate such pairs as SLOW-*Fast* and BLUE-*Yellow*.[2] Apart from the implicit assumptions made by collapsing disparate stimulus and response terms together, the possibility is ignored that certain reciprocal pairs, for example, WORK-*Play*, PLAY-*Work*, bear no simple consistent relation to one another.

For this reason, it seems more sensible to simplify the problem by dealing with the distribution of responses with stimuli held constant or with the distribution of stimuli for any given response. The distributions to different stimuli or the stimuli for pairs of responses may then be compared. This comparison should provide the first step in understanding how associations are related to one another.

RELATIONSHIPS AMONG ASSOCIATES

For the past several years, the present author has been studying the extent to which the relations among associations determine such aspects of immediate free recall as clustering, importations into recall, etc. (Deese, 1959a, 1959b, 1960). This work had its origin in the studies of associative clustering (Jenkins & Russell, 1952) and category clustering (Bousfield, 1953). The basic method has been to relate an index based upon the average frequency with which all words presented for recall tend to elicit one another as free associates with the number of words correctly recalled, the number of importations, etc. This index is a powerful predictor of a number of characteristics of free recall. Free recall itself is a representative verbal process, and thus it is evident that the associative interrelations among words are powerful determiners of the pattern and frequency with which words are emitted.

---

[2] In this paper, free association stimuli will be given in small capitals and responses in italics.

The interword associative index is computed by tabulating the frequencies with which each word in a particular collection (usually a list of words to be presented for recall) occurs as a response to all of the other words as free association stimuli. Casual inspection of the matrix that provides the basis for such a tabulation gives striking confirmation of what every investigator who has studied free association has felt to be the case, namely that associations exist in well organized and in some instances tightly organized networks. The networks are so apparent that they turn up even in random collections of words, if the words are high in frequency of usage (Deese, 1960).

The networks that appear upon such a tabulation provide only part of the picture, however, because the words that appear as responses are artificially restricted to words which appear as stimuli. When examined, the original association frequencies show that stimuli which do not elicit one another as responses sometimes have in common a great many responses. An example, picked more or less at random, is the pair PIANO and SYMPHONY. In a sample of Johns Hopkins University undergraduates, neither one elicited the other, but both elicited *Note, Song, Sound, Noise, Music*, and *Orchestra* in varying frequencies.

Thus, the relationship between free association stimuli should be determined from *all* of the responses that stimuli have in common. If the associative meaning of any stimulus is given by the distribution of responses to that stimulus, then two stimuli may be said to have the same associative meaning when the distribution of associates to them is *identical*. Two stimuli overlap or resemble one another in associative meaning to the extent that they have the same distribution of associates.

One may also deal with the associative distribution [3] of the responses themselves. The associative distribution of a response is the collection of stimuli to which it occurs. This relation is less interesting than that between responses to stimuli for a number of reasons. It does not, of course, specify an associative concept as does the distribution of responses to a stimulus, and it is subject to serious sampling limitations. One is fairly sure to pick up any relations of statistical significance among stimuli by a modest sample of responses. With the distribution of stimuli problem, however, there is no way to insure that one has sampled all stimuli that are *high* frequency producers of a particular response short of exhausting the language. Therefore, except insofar as the distribution of stimuli problem is related to the problem of associative meaning, it will be ignored in the balance of this paper.

SPECIFICATION OF RELATIONSHIPS

The idea of associative meaning presented above implies that one describes and classifies associations purely in terms of the relations among the responses to different stimuli. There are several ways in which such relations may be specified. One may count the number of associations in common in any fixed size sample, or one may weigh the common associations by their frequency. For prediction of effects in verbal behavior, the latter procedure is probably more precise, and it is the approach taken in the present paper. Also, one may wish to describe the *ways* in which stimuli resemble one another in associative meaning. This implies the study of a collection of relations between a number of stimuli. Factor analysis suggests itself as a way of reducing the structure within such a collection to manageable proportions. In any event, the heuristic value of the concept of associative meaning will depend upon ways of specifying the relations between free association stimuli. The purpose of this section is to provide a basic method for measuring the relationship.

---

[3] The word "distribution" is used here in analogy to its linguistic, not its statistical sense. That is to say it indicates the number of situations to which a particular word occurs as a response.

There are several difficulties that stand between the definition of associative meaning and the specification of particular associative meanings by a study of the relationships among free associations. The first of those is created by the technique of collecting free associations, which almost completely eliminates the occurrence of the stimulus word as a response to itself. This problem must be dealt with first.

The first step in an analysis of associations is to tabulate for each stimulus word in a collection the responses which the word has in common with any other stimulus words in the collection. Since, by instructions as well as by technique a word usually does not occur as a response to itself in free association, the cells in such a tabulation in which the same word as response and stimulus intersect will be empty. This creates intuitive difficulties if one defines the relation of associative meaning between words as the common distribution of responses, because pairs of words often appear to be intuitively related when they have no *overt* responses in common. Upon close examination all such intuitive cases turn out to be cases in which the responses to one word are the stimuli to the other. An example in the Hopkins data is the SOFT-LOUD pair. Both of these words elicit one another, but they have no other responses in common. Thus, they could have no associative meaning in common if the possibility of a word occurring as a response to itself is denied.

Fortunately, a way out of this paradoxical situation already exists in a technique independently designed for other purposes by Jenkins and Cofer (1957) and by Bousfield, Cohen, and Whitmarsh (1958). The technique consists of assuming that a word elicits itself implicitly as an associative 100% of the time. This assumed implicit response is called by Bousfield, Cohen, and Whitmarsh the "representational" response. The implication of the name is that this response identifies or represents the symbolic stimulus presented to the individual.

Both Jenkins and Cofer (who do not explicitly identify the representational response) and Bousfield, Cohen, and Whitmarsh implicitly make the assumption that the representational response is always the same word as the stimulus presented. This assumption is very nearly correct for words of common cultural use, but it must surely be incorrect for unusual words. Many Hopkins undergraduates, for example, respond to the words ABBESS and ABYSS in precisely the same way, which suggests that the representational response for one of them is not the dictionary equivalent of the word. Nevertheless, this assumption works with only trivial exceptions for moderately sophisticated individuals and words of ordinary frequency of usage. It is, therefore, adopted here.

With this assumption it is possible to prepare a table like Table 1. This table shows the frequency of responses common to any two stimuli, with the frequency of the stimulus word as a response to itself entered with 100% frequency. Since the *N* for this table is 50, the value entered in the stimulus-response intersection cells is 50. Note that it is not necessary for all the stimulus-response intersections to appear. If a particular stimulus word never appears as a response to the other stimuli, it will not appear as a common response. Such instances are rare, however, when the collection of stimuli is made on some basis that predicts related associative meaning.

The second problem is to find an appropriate measure of the relation between the responses common to different stimuli. This is the problem of finding a measure of the relations between the columns in Table 1. The most obvious solution is to use product-moment correlations as the measure of relations between columns. There are, however, a number of objections to such a solution.

The most important objection, determined in part by the fact that the technique of free association limits each subject to one response, is that the product-moment correlation would yield a measure of the correlation of *distributions* of associations rather than the extent to which frequencies are in common. Often, responses that are high in frequency to one stimulus will be low in frequency to another. This happens very often because of the typically steep rank-frequency distribution of free association responses to any given stimulus. Because the forms of the distributions of *common* responses may be inverse to one another, the obtained product-moment correlations of the frequencies will often be negative. Thus, a table of product-moment correlations obtained from Table 1 would yield a few zero correlations, a few high positive correlations and a large number of low negative correlations.

Since it is not the relation between the *forms* of the distributions of responses but the relations between the frequencies themselves that is interesting, a direct measure of overlapping frequency is appropriate. To measure the commonness of associative meaning, we may take the ratio of the sum of the overlapping frequencies to any pair of stimuli to the maximum possible sum. Thus, if we turn to the first two columns in Table 1, which are the responses to MOTH and INSECT respectively, we see that the response frequencies in common sum to 12. The total possible frequency in common is 100; this yields a relative common frequency of .12. This number is entered into the appropriate cell in Table 2.

A simple interpretation of such a measure demands some assumptions which are not strictly correct. In other words, for most problems, this measure can be regarded as an approximation. It supposes that the distribution of responses taken one per individual per stimulus word from a homogenous population is characteristic of the distribution of associations at different times in any one representative individual from that population. Thus, the assumption is made that the distribution of responses obtained from a sample of individuals is the same as the distribution of associations in any one individual.

This is an assumption shared by most students of free associations in population samples (and one only occasionally explicitly stated). An experimental comparison of repeated testing vs. single testing is a badly needed addition to the association literature, and it may well lead to a more satisfactory way of counting instances of commonness of association.

This measure of the relation in associative meaning between two words can

TABLE 1

RAW FREQUENCIES OF ASSOCIATES IN COMMON TO 19 WORDS BASED ON RESPONSES OF 50 SUBJECTS

Responses	Stimulus words[a]																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 *Moth*	50	2						1	8										7
2 *Insect*	1	50			3				3										6
3 *Wings*	2		50	4															5
4 *Bird*			25	50	4			1	2			9					2		4
5 *Fly*	10	9	12	15	50			2				1					1		4
6 *Yellow*						50	2				1			4					3
7 *Flower*						2	50				1	2		1	10		1	2	2
8 *Bug*		24			4			50	5			1							
9 *Cocoon*									50										2
10 *Color*						5				50	6								
11 *Blue*				1	1	2	2			8	50			1		40			
12 *Bees*		1					2	2				50							
13 *Summer*	2								1			1	50	1	1				
14 *Sunshine*													1	50				12	
15 *Garden*							6								50				
16 *Sky*					1						6			1		50			
17 *Nature*																	50	1	
18 *Spring*													3					50	
19 *Butterfly*	1								8										50
20 *Light*	4					1								4		1		1	
21 *Pretty*							3												2
22 *Ant*		3			1			5											
23 *Bright*										1				4					
24 *Airplane*			4		1														
25 *Feather*			2	3															
26 *Flight*			1	2															
27 *Tree*				2	1									1			6		
28 *Plane*			2		5														
29 *Red*						6	1			16	13								
30 *White*						1				5	2								
31 *Green*						5	2				4		1		3		2		
32 *Sun*						2													
33 *Beetle*		1						1											
34 *Spider*		1						1											
35 *Gold*						1					1		1						
36 *Black*		1				1				8	2								
37 *Winter*													17					4	
38 *Warm*													3	8				1	
39 *Plant*							2								5		1		
40 *Gray*						1				1	2								
41 *Brown*						1				1									
42 *Vacation*													2					1	

[a] The numbers of the stimulus words correspond to the first 19 response words.

TABLE 2

OVERLAP COEFFICIENTS FOR COMMON ASSOCIATES BETWEEN THE 19 WORDS IN TABLE 1
(Decimals Omitted)

Stimulus words	Stimulus words																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 MOTH	100	12	12	12	11	02	00	05	11	00	00	02	02	05	01	01	01	01	15
2 INSECT		100	09	09	17	01	01	33	10	01	01	03	00	00	00	00	01	00	12
3 WING			100	44	19	00	00	03	02	00	00	10	00	00	00	00	03	00	13
4 BIRD				100	21	01	00	03	02	01	01	10	00	01	00	01	05	00	12
5 FLY					100	01	01	08	06	01	02	06	00	03	00	02	04	00	11
6 YELLOW						100	07	00	00	17	23	02	02	07	05	02	04	03	05
7 FLOWER							100	02	00	03	07	02	01	06	18	02	06	02	06
8 BUG								100	07	00	00	05	00	00	00	00	02	00	04
9 COCOON									100	00	00	04	01	01	01	00	02	00	22
10 COLOR										100	32	00	00	02	00	08	00	00	00
11 BLUE											100	01	02	04	04	46	03	02	02
12 BEES												100	01	02	03	00	04	02	07
13 SUMMER													100	05	02	00	01	10	00
14 SUNSHINE														100	02	03	02	15	04
15 GARDEN															100	00	04	02	02
16 SKY																100	00	01	00
17 NATURE																	100	02	03
18 SPRING																		100	02
19 BUTTERFLY																			100

vary from .00 to 1.00. The latter value would happen, however, only if the two stimuli produce completely reciprocal responding. Thus, complete identity in associative meaning can occur by this measure *only* if two stimulus words *always* elicit each other and *no other* words. This is a consequence of the assumption of the frequency of the representational response. Thus, if a pair of stimuli had distributions of responses exactly alike except that they never elicited each other, the maximum relative frequency in common would be .5. Such a result appears at first to be a very undesirable property of the measure of relation in associative meaning. However, it turns out, as will be pointed out later, that this property is extremely important in generating some useful consequences of the concept of associative meaning.

The relative overlap coefficients for all pairs of stimuli within a set produces a matrix of the sort found in Table 2. The only problem remaining is to fill in the principal diagonal, or the communalities in an ordinary table of correlations. In keeping with the assumptions made above about the assessment of associative meaning, the proper value to enter into these cells is 1.0, since any word always has the same distribution of response frequencies as itself. This is like the assumption introduced by Osgood and Suci (1952) into their factorial portrayal of the semantic differential.

ANALYSIS OF ASSOCIATIVE MEANING

Before examining some of the implications of this notion of associative meaning, it would be helpful to look at some sample data showing some of the relations that can exist between the associative meanings of different words.

It would be possible, of course, to apply the present analysis to random collections of words, but because the overlap in associative meaning between words taken at random is not high, such an application would not be rewarding at this stage. In general, an analysis of associative meaning would be conducted on a set of stimulus words selected because we would be interested in discovering their relationships and because we would suspect, from prior information, that they were associatively related.

The data presented here illustrating some of the possible relations in associative meaning are not psychologically interesting, with one exception. They were, however, selected in such a way as to insure some degree of related associative meaning; thus they illustrate fairly well some of the analytic possibilities.

The sets of stimulus words in Tables 3, 4, and 5 have internal structure for the simple reason that the stimulus words themselves are all *responses* to one of the stimulus words by the Minnesota norms (Russell & Jenkins, 1954). For example, the words in Table 3 include BUTTERFLY and words which are responses to BUTTERFLY in the Minnesota norms.

These data are based upon samples of 50 subjects each, which are a bare minimum for any kind of reliable analysis and too small for anything but the roughest kind of prediction. They serve, however, to illustrate how sets of words are related to one another and what the factor structure of associations is like. The subjects from which these associations were obtained are undergraduate students (male) at the Johns Hopkins University.

The basic data from which the factors in Table 3 are derived are in Table 2. These are relative common frequencies of responses to 19 words in the BUTTERFLY set. These relative frequencies were factored by the centroid method, and after a very few rotations (pairs of factors were rotated through approximately 45 degrees) the factor loadings presented in Table 3 were obtained. The factor loadings in Table 4, based upon stimulus words obtained from MUSIC, and Table 5, based upon stimulus words obtained from SLOW were likewise arrived at by centroid analysis and very simple rotation.

Since the major emphasis in these sample data is on the factor presentation, a word needs to be said about factor analysis in this context. Factors, of course, depend upon the variables used to obtain them. In testing theory this is a sometimes overlooked restriction if for no other reason than that test theorists are often convinced that the batteries they factor are in some sense "representative" of all the set of abilities they wish to factor. In the analysis of words, however, this is not the case, so we are forced to an awareness of the limitations of factor extraction. The words that go into any matrix are only a small part of those that might possibly have been used. Thus, the results of the factor analysis are properties of the limited set, and we cannot talk about the factor loadings for any one word, for example, as if these exhausted all of the associative structure of that word. As it turns out, this is a very useful property of the analysis of associative relations among words, but it tends to make trivial collections of words, such as those in Tables 3, 4, and 5 less interesting in a factor analysis than would be collections of words with some suspected important structure to be uncovered.

The first thing to notice about the factor loadings is the convenient simple structure in the factors. In Table 3, approximately half of the words have nearly zero loadings on Factor I, while the other half of the words have positive loadings of about the same magnitude. Those words on Factor I that have approximately zero loadings have positive loadings on Factor II, and the words with positive loadings on I have nearly zero loadings on II. Thus, the first two factors effectively sort the stimulus words into two classes.

This kind of initial classification is a property of the kind of distribution of common frequencies found in Table 2, and since this distribution is typical of all of the 15 or so matrices thus far examined by the author, it will probably occur in all centroid analyses of words. The same initial classification occurs in Tables 4 and 5, and it also occurs in Table 6 for which the stimulus words were not selected for associative relations but because they centered about a particular attitude or value system.

An inspection of the loadings of Factor I in Table 3 shows that it occurs in words having to do with animate creation (BEES, FLY, BUG, WING, BIRD, etc.) whereas the words with loadings on Factor II are those words in the set that do not have to do with animate things (SKY, YELLOW, NATURE, etc.). It would be a mistake, however, to fall into the practice of naming factors derived from associative distributions, for inevitably we fall into the attempt to impose a *logical* structure on the associations, and one thing that is clear is that associative meaning is not logical. Nevertheless, it is clear that the first two factors have split the stimulus words into the most *obvious* and *basic* structure organizing them.

TABLE 3

ROTATED CENTROID FACTOR LOADINGS OF STIMULUS OVERLAP COEFFICIENTS PRESENTED IN TABLE 2
(Decimals Omitted)

Words	Factors					
	I	II	III	IV	V	VI
Moth	44	03	−27	−01	−03	−32
Insect	50	01	−33	01	−34	11
Wing	52	01	45	01	29	−07
Bird	52	02	46	01	29	−07
Fly	48	03	32	01	−28	−03
Yellow	01	44	−03	34	−32	−02
Flower	01	39	−03	−32	03	44
Bug	41	01	−34	00	−14	37
Cocoon	40	01	−35	00	25	02
Color	−02	42	−04	44	04	−04
Blue	−02	57	−04	52	23	−04
Bees	36	04	34	−02	−30	00
Summer	−01	31	−03	−34	−02	−34
Sunshine	02	37	−04	−33	−03	−35
Garden	00	35	−02	−34	−03	44
Sky	−01	41	−03	43	38	−07
Nature	04	31	29	−02	01	34
Spring	−01	35	−03	−37	−02	−36
Butterfly	48	06	−29	−01	26	01

Factor III in Table 3 shows zero loadings on the nonanimate words again (except for NATURE, which apparently comes close to straddling the basic division of words in Table 3), and it splits the animate words into a bipolar factor. The positive loadings are on WING, BIRDS, BEES, FLY and the negative loadings on BUG, COCOON, MOTH, BUTTERFLY (which probably occurs here because of its strong relation to BUG and COCOON). Factor IV makes a bipolar split of the nonanimate words. SUMMER, SUNSHINE, GARDEN, FLOWER, and SPRING go together while BLUE, SKY, YELLOW, and COLOR go together.

It is instructive to follow the factor profiles for pairs of words within a set. This allows us one of several ways of preparing an associative thesaurus, in this case a pair thesaurus. Thus, we may say that BLUE and YELLOW are both alike within this set in that they share common loadings with SKY, SUMMER, SUNSHINE, COLOR, etc. On one factor, however, they diverge; this provides the critical difference between BLUE and YELLOW within this set of words. BLUE goes with SKY, BUTTERFLY, WING, and BIRD, while YELLOW goes with INSECT, FLY, BUG, and BEES. This, of course, does not exhaust the contrasts in associative meaning between BLUE and YELLOW, but it does suggest one way in which they differ.

Table 4 presents the first six factors from a centroid analysis of the matrix of stimulus words which are themselves responses to MUSIC. Again, the first two factors yield a basic split. Factor I has loadings on TONE, INSTRUMENT,

TABLE 4

ROTATED CENTROID FACTOR LOADINGS OF STIMULUS OVERLAP COEFFICIENTS FOR 17 WORDS DERIVED FROM MUSIC (Decimals Omitted)

Words	Factors					
	I	II	III	IV	V	VI
Tone	19	43	−03	40	02	53
Instrument	43	−01	33	01	−31	01
Symphony	49	06	34	01	43	00
Sing	41	−01	−53	02	−36	−01
Note	43	01	−33	03	37	−01
Song	55	−01	−47	00	−31	01
Sound	04	59	00	−26	41	07
Piano	52	02	37	01	−24	00
Noise	06	59	01	06	02	31
Band	51	08	38	−04	−24	−03
Horn	37	−09	00	−32	00	−43
Loud	00	48	−01	−42	−01	−29
Hear	06	60	−01	49	01	−18
Opera	51	03	−33	01	32	00
Ear	00	41	−01	52	−02	−36
Music	58	08	25	01	37	02
Soft	07	23	−02	−01	00	07

SYMPHONY, SING, NOTE, SONG, PIANO, BAND, HORN, OPERA, and MUSIC, while Factor II has loadings on SOUND, NOISE, LOUD, EAR, and HEAR (this is a less clean split than the previous table because of the small loading of TONE on Factor I). As with the previous table, Factors III through VI are bipolar. Again, interesting contrasts may be seen in factor profiles. OPERA and SYMPHONY, for example, have parallel profiles except on Factor III where they are bipolar opposites. Here, OPERA goes with SONG, NOTE, and SING, while SYMPHONY goes with INSTRUMENT, PIANO, BAND, and MUSIC.

Table 5 exhibits exactly the same sort of pattern with the words centering around SLOW and so will not be commented on further. This table illustrates an effect to be found in associations in adults, the close relations of opposites. The pair SLOW-FAST yield an overlap index of .71 for the Hopkins population, the highest for any of the words in Tables 3, 4, and 5. For the six factors extracted, the factor profiles for these two words are practically identical. This means that there is not enough contrast provided within the matrix of words to separate these words within the arbitrary limit of six factors. This is in contrast with the pair SOFT-LOUD, which separates almost immediately in the MUSIC table. The presence of NOISE, HEAR, HORN, etc., managed to pull SOFT and LOUD apart by providing some associative meaning overlapping with LOUD but not with SOFT.

Because of the close relation between SLOW and FAST, it may be very difficult to provide sets which will contrast these in associative meaning. They are, as are most of the high frequency opposites, adjectives, and they are descriptive of a common state modified only in a kind of secondary way. That

TABLE 5

ROTATED CENTROID LOADING OF STIMULUS OVERLAP COEFFICIENTS OF 16 WORDS DERIVED FROM SLOW (Decimals Omitted)

Words	Factors					
	I	II	III	IV	V	VI
Slow	67	00	39	−01	20	−01
Walk	41	02	−55	00	−38	03
Speed	49	09	−33	07	47	−04
Quick	61	00	25	−02	−07	−28
Lazy	04	35	05	−38	−05	−36
Drive	10	39	−33	02	00	04
Skid	05	40	−01	36	35	−01
Run	54	01	−52	−01	31	−01
Work	03	33	36	−01	05	−36
Fast	73	01	38	−02	20	−02
Down	01	31	−01	−36	−07	−37
Stop	01	37	05	37	−09	36
Snail	43	02	37	02	−10	33
Sign	00	36	07	07	02	−44
Poke	02	34	05	−37	−05	35
Traffic	04	36	06	33	−04	−36

is to say, both words imply motion; they both describe the rapidity of motion, and they contrast only in the kind of rapidity. The way in which SOFT and LOUD are alike, however, provides only part of the distribution for SOFT, since it exists also in a close relation with the word HARD. HARD and LOUD are related through SOFT but otherwise independent.

Table 6 illustrates the factor analysis of a set of stimulus words which are psychologically more interesting. These words are taken from a larger study, to be presented elsewhere, which is concerned with the factor structure of associative meaning of critical words for subjects scoring high and subjects scoring low in each of the value scales from the Allport, Vernon, and Lindzey (1960) study of values inventory. The critical words for each value scale were obtained from the actual items scored for that value scale; thus the words in Table 6 are from the Religious Value items.

TABLE 6

Rotated Centroid Loadings of Stimulus Overlap Coefficients for 15 Words Centering in Religious Value

Words	Factors					
	I	II	III	IV	V	VI
Sermon	03	43	38	−10	04	33
Clergyman	11	42	05	−47	00	26
Religion	54	03	−02	−34	−03	−36
Service	−01	40	−02	−43	37	−16
Worship	57	04	33	−05	−30	−02
Reverence	55	04	36	02	27	−03
Prayer	54	03	34	−03	−01	−34
Soul	06	38	−02	34	03	39
Spirit	39	00	−37	05	39	05
Divine	50	00	05	35	34	04
Faith	52	08	−36	−04	−35	01
Inspire	03	40	−01	33	−01	−41
Devotion	49	10	−33	02	00	34
Love	03	38	00	35	36	05
Hope	06	37	−41	−01	−04	−45

Completed data are available at present only on high scorers (the distribution of scores in the Hopkins population is skewed). Thus, the factors in Table 6 are based upon 50 Hopkins undergraduates who scored at or above one probable error in the published norms for men on religious value. The first thing to notice about Table 6 is that it is very much the same as Tables 3, 4, and 5. The initial classification in this case sorts the words into a group including DIVINE, SPIRIT, PRAYER, WORSHIP, REVERENCE, RELIGION, FAITH, and DEVOTION and into a group including HOPE, SOUL, LOVE, INSPIRE, SERVICE, SERMON, and CLERGYMAN. One is almost tempted to say that the classification divides the words into an institutional, theological class and a human individual class.

Implications of Associative Meaning

By now, some readers will have recognized the notions presented here as extensions of those stated by Bousfield, Cohen, and Whitmarsh (1958), and since these authors regard their work as an extension of Noble's (1952), by implication the present work is also.

Thus far we have asserted that the distribution of responses to any free association stimulus[4] forms the associative meaning of the stimulus word and thus an associative concept named by that word. If the associative meaning of a word (as opposed to any particular stimulus-response pair) is to have any use, it must predict something about verbal behavior. In this section we will discuss some of the characteristics of verbal behavior that may be predictable from the associative meanings of words and from collections of associative meanings.

---

[4] Free association stimuli are not limited to words, of course, and the ideas presented here can be in theory, extended to nonverbal stimuli, but the lack of a clear criterion for the representational response may produce difficulties in practice.

Associative meaning will not predict the tendency of words to elicit one another; that is predicted by the free association frequencies of particular pairs or by some general measure of pair relation (such as the interword associative index). Associative meaning, in general, should predict the words that will occur in the verbal environment of a particular word. This statement entails the assumption that words are used, subject to certain other constraints, in particular environments because of the distribution of associations they possess. Thus, the distribution of associates takes on the character of mediation determining the use of particular words.

Thus, if a particular word appears in some verbal environment, a word close to it in associative meaning should appear in the same environment. Generally, the closer the relation, the higher the probability of the two words appearing in the same environment. The words may appear in the same environment in two ways: (*a*) as substitutes for one another, or (*b*) as part of one another's environment. It is not a part of associative meaning to predict which of these two ways will appear; though such a prediction should arise out of theory about the origins of associations.

The first kind of environmental relation (that of substitution) is illustrated quite clearly by the highly related SLOW-FAST pair. The environment that produces *Slow,* all other things equal, is very likely to produce *Fast* also, but not nearly so likely to produce dictionary or thesaurus equivalents of *Fast.* Thus, we are likely to say "this is a slow train" or "this is a fast train," but we are unlikely to say "this is a rapid train" or "this is a sluggish train."

The second kind of environmental relation (that of coordination) is illustrated by the SKY-BLUE pair. Given the frame, "he ——— into the ——— sky," a highly likely substitution in the second blank is "blue." All words in English seem to be able to produce both kinds of environmental relation, but for many words one or the other type predominates. Thus, while "the blue" may appear as a substitution for "sky," such useage is poetic and rare.

One problem is produced by the apparent assymetry of substitution in particular frames. It is possible, indeed probable, that the frame, "he ——— into the blue ———," will not produce "sky" with the same frequency that "blue" is produced in its complementary frame. Particular test frames have an influence, and it is possible that substitution frequencies may be symmetrical when averaged over a representative group of test frames. Such symmetry is a property of the definition of associative meaning, though the property cannot be properly tested, since any failure of symmetry may easily be attributed to a lack of adequate sampling.

Earlier it was pointed out that one of the consequences of the assumption of the representational response was that perfect agreement in associative meaning between two words would occur only if they were perfectly reciprocal, that is to say, if they were the only associates of each other. While this seemed at first to be an undesirable property, it does conform with the limiting case of substitution. For, if two words are perfectly correlated in associative meaning, then they ought to be the only substitutes for one another in the substitution relation. Notice that this situation would not theoretically prevent these words from being part of the associative meaning of a third word. Thus Word A could elicit only B as a free associate and B elicit A, which would make them overlap completely in associative meaning, but both A and B could be free association responses to a third Word C. Obviously this situation would not happen in any real language, but an approximation to it frequently occurs. The general rule for such a case is that the reciprocal pair of words are both of high frequency of usage, while the third word, which elicits them both but never occurs as a response to either, is a word of very low frequency of usage. An example in the Hopkins data is the overlapped pair WOMAN and GIRL, neither of which elicit *Virginal,* but both of which are elicited as responses by VIRGINAL.

The use of test frames to validate the notion of associative meaning will be difficult in practice, since the influence of grammatical constraints, etc., needs to be removed. Furthermore, we do not know enough about the problem of compounding to know whether or not compound effects can be derived from the concept of associative meaning itself. Nevertheless, whatever other interesting effects may be achieved, the ultimate validation of the concept of associative meaning must come from its ability to predict substitutions in the flow of language.

The most important use of the notion of associative meaning is to study the structure of sets of words having intrinsic psychological interest. The number of such sets is very large, and indeed is only limited by the ideas psychologists have about the influence of psychological variables upon the use of words.

An obvious application is to any collection of words as metaphorical symbols. Those words (or concepts) which by psychoanalytic theory, for example, are supposed to symbolize the phallus, the female figure, hostility, etc., ought to have their common symbolic character revealed in the interrelations and factor structure of their free associates. This is not to say that phallic symbol words ought to reveal a common core of phallic meaning (by dictionary criteria) but that they ought to converge on some common factor. In a word, no matter how remote their dictionary meanings and references in the real world, they should reveal a common associative meaning. If they do not, it is hard to see how the particular psychoanalytic symbol investigated could have any general validity for the population tested.

Currently, the present author is studying the structure of associative meaning within a set of words that were selected from Rorschach protocols as indicators of aggression. No detailed account of these data can be given at present, but a striking qualitative aspect of them is the extent to which the overlap in associative meaning among them is mediated by a small number of response words, many of which appear out of context to be relatively neutral (*Stick,* for example).

The technique may be used also to study the verbal organization associated with attitudes. Table 6 presents the factor structure of associations to words centering about religious values for individuals who score high in religious value. A complete set of data is not yet available for low scorers, but it is almost certain that the most important difference between the high and low scorers will be in the extent to which the associations to these words are well organized. The factor loadings for low scorers will almost certainly be lower; indeed, many of the stimulus words (SERVICE, LOVE, HOPE, SPIRIT) will all but disappear from the factor structure, and it is very likely that the factor structure of the remaining words will be very different.

ORIGIN OF ASSOCIATIONS AND SUMMARY

The last question we face concerns the origin of associations. The view of Locke that attributed them to the accidental contingencies of nature has led to a kind of mosaic theory of mind, a theory that has scarcely ever had a serious rival on its own ground, though it has often been challenged.

No real substitute for the contiguity-similarity postulate is offered here. The highly organized economy of associative meaning has impressed the author, however, and it was a belief that the human mind derived associations from categories of its own that sent him on the search for a technique by which to study associative meaning. Thus, the least that can be offered is the suggestion that associations derive in whole or part from the structures or categories of the human mind. This belief is probably the mover of attempts, such as Woodworth's, to classify associations. Such attempts are fruitless, however, for if there are categories in association, they are categories of association, not categories of subordination, coordination, etc. This is admittedly the belief that motivated the present work.

Thus, the implicit distributions of responses that define associative meaning may exhibit patterns of overlap because they are derived from simple structures. An adult may use the word BUTTERFLY to correspond to a class of living creatures that fly and are bugs. This doesn't however, exhaust the associative meaning of BUTTERFLY, for most flying bugs are dirty, unpleasant and YELLOW, while BUTTERFLIES are pretty, like BIRDS and flutter in the BLUE SKY. Thus, another associative category intrudes on the meaning of BUTTERLLY and makes it an associative concept not quite the same as any other. Contrary to zoology, associative BUTTERFLIES are as closely related to the BIRDS as to the MOTHS.

Large collections of words ought to enable us to discover the categories of association, if they exist, and how they change with age, if they do. In any event, an alternative to the contiguity-similarity tradition in association theory can come about only by a study of the associations themselves, and the purpose of this paper is to present a somewhat different approach to the study of such associations. That approach consists in the assumption that the distribution of associates to any word provides the associative meaning of that word and the techniques necessary to apply this assumption to data.

REFERENCES

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *A study of value.* (3rd ed.) Boston: Houghton Mifflin, 1960.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.,* 1953, **49**, 229–240.

BOUSFIELD, W. A., COHEN, B. H., & WHITMARSH, G. A. Verbal generalization: A theoretical rationale and an experimental technique. Technical Report No. 23, 1958, Contract NONr onr 631(00).

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.,* 1959, **5**, 305–312. (a)

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.,* 1959, **58**, 17–22. (b)

DEESE, J. Frequency of usage and number of words in free recall: The role of association. *Psychol. Rep.,* 1960, **7**, 337–344.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.,* 1952, **47**, 818–821.

JENKINS, PATRICIA M., & COFER, C. N. An exploratory study of discrete free associations to compound verbal stimuli. *Psychol. Rep.,* 1957, **3**, 599–602.

Karwoski, T. F., & Berthold, F. Psychological studies in semantics: II. Reliability of free association tests. *J. soc. Psychol.,* 1945, **22**, 87–102.

Noble, C. E. An analysis of meaning. *Psychol. Rev.,* 1952, **59**, 421–430.

Osgood, C. E., & Suci, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.,* 1952, **49**, 251–262.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for response to 100 words from the Kent-Rosanoff Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr 66216.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received October 6, 1960)

Psychological Review
1962, Vol. 69, No. 3, 176–186

CONTIGUITY AND REINFORCEMENT IN RELATION TO CS–UCS INTERVALS IN CLASSICAL AVERSIVE CONDITIONING[1]

JOAN E. JONES[2]
University of Sydney

It is generally accepted that, in classical aversive conditioning, the best results are obtained if there is a temporal interval of approximately 450 milliseconds between the presentation of the conditioned stimulus (CS) and the uncondtiioned stimulus (UCS). This observation creates a problem for theorists who adopt the common sense notion of contiguity as a basis for the association of events in classical conditioning. Such theorists formulate the problem as one of explaining why separation of CS and UCS is superior to contiguity and of why an interval of 450 milliseconds is better than others. Because the problem has been formulated in this way, the theoretical constructs proposed are rather arbitrary and have not been subjected to experimental tests. For example, Hull (1952) proposes a theoretical construct which bridges the temporal gap between the stimulation of a receptor by the CS and the evocation of the unconditioned response (UCR), namely the molar stimulus trace. The same function is performed by Guthrie's (1935) concept of proprioceptive stimulation, though he does not explain why certain trace conditions are actually superior to the simultaneous presentation of CS and UCS. Cognitive theorists invoke the perceptual principles governing grouping, particularly temporal proximity (e.g., Tolman, 1949). Hilgard and Marquis (1940) interpret the classical situation in terms of reinforcement rather than of the contiguity of associated events, basing their interpretation on the observation that the most favorable interval appears to be somewhat greater than the latency of the conditioned response (CR).

Actually there is less agreement in the literature on optimum intervals than the usual textbook statement implies. Experimental evidence on the effect of the CS–UCS interval in conditioning human subjects is summarized in Table 1. In the experiments cited the classical defense method was used with the exception of Wolfle's where it appears that avoidance was possible. In considering Table 1 it should be noted that each of the studies reported has observed the acquisition of only one response. In comparing any two studies it must be recognized that many experimental conditions are varied even where different experimenters have employed the "same" response. The most important of these is the experimenter's original choice of intervals which may impose severe limitation on the optimum obtained. The range and the spacing of the chosen values have contributed to both artifactual similarities and dissimilarities in optima reported by different investigators. From Table

---

[1] Based on a thesis submitted in partial fulfilment of the requirements for the PhD degree, University of Sydney, 1959, supervised by R. A. Champion. The author expresses appreciation to R. B. Bromiley and M. Humphries for their helpful criticisms of a draft of this article.

[2] Now at Defence Research Medical Laboratories, Toronto.

TABLE 1

EXPERIMENTAL STUDIES OF CS–UCS INTERVAL EFFECTS

Response	Author	Intervals sampled[a]	Optimum[a]
Eyeblink	McAllister (1953)	100, 250, 450, 700, 2500	250
	Bernstein (1934)	−900, −500, 100, 200, 250, 300, 500, 1000, 1480	300
	Hansche and Grant (1960)	190, 390, 590, 790	390 (training and extinction)
	Kimble (1947)	100, 200, 225, 250, 300, 400	400
	Reynolds (1945)	250, 450, 1150, 2250	450
	Myers (1950)	500, 1000, 1500, 2000	500
Finger withdrawal	Wolfle (1932)	−2000, −1000, −600, −200, 0, 200, 300, 400, 600, 1000, 2000, 3000	300 (avoidance possible)
	Wolfle (1930)	−500, −250, 0, 250, 500, 750, 1000, 1250, 1500	500 (avoidance possible)
	Spooner and Kellogg (1947)	−500, −250, 0, 500, 1000, 1500	500
Pupillary dilation	Gerall and Woodward (1958)	250, 500, 1500, 2500	1500
GSR	Moeller (1954)	250, 450, 1000, 2500	450
	White and Schlosberg (1952)	0, 250, 500, 1000, 2000, 4000	500 (extinction)
	Wickens, Gehman, and Sullivan (1959)	500, 600, 700, 800, 920, 1040, 1550, 2500, 4600	500 & 1500 (two optima)
	Jones (1959)	240, 550, 740, 1060, 1575	550 (training) 1060 (extinction)
	Jones (1959 unpublished)	500, 1000	1000
	Bierbaum (1959)	−3000, 0, 500, 3000, 5000	3000

[a] in milliseconds

1 it will be seen that all the studies employing short latency responses (eyeblink and finger withdrawal) report optima of 500 milliseconds, or less, and as short as 250 milliseconds, while the studies in which responses of longer latency have been used (pupillary dilation and GSR) yield optima of 450 milliseconds or more, even as long as 3000 milliseconds. The difference is in the direction expected by Hilgard and Marquis but the optima reported in five of the six GSR studies are well short of the latency of that response.

This paper presents an interpretation of the effect of the CS–UCS interval in terms of the contiguity of the CS and the UCR and of reinforcement. This interpretation does not begin with the assumption of an invariant optimum interval but accepts the empirically demonstrated fact that no single optimum exists. The propositions to be outlined differ from existing two-process theories of conditioning. Whereas Skinner (1935) and Schlosberg (1937) invoke one principle (contiguity) in classical conditioning and another (reinforcement) in instrumental conditioning, the present theory invokes the action of both principles in classical conditioning. It differs from Mowrer's (1947) two-factor theory of classical conditioning in that it contains no reference to intervening emotional reactions and no distinction is made between the principles governing the conditioning of responses on the basis of their neurophysiological characteristics.

A TWO-PRINCIPLE THEORY OF CLASSICAL CONDITIONING

When the CS–UCS interval is varied two temporal relationships are affected. One is the interval between the CS and the UCR, pertaining to the principle of S–R contiguity. The other is the interval between the CR and the UCS, which defines the delay of reinforcement. These changes are effected at the same time but may differ in extent and/or direction.

The following assumptions are made about the operation of S–R contiguity and of reinforcement in classical conditioning:

S–R Contiguity. The optimum condition for the establishment of a new S–R connection is that the CS and the UCR be contiguous in time. The greater the degree of temporal separation of these events (in either direction) the weaker the resulting connection.

Reinforcement. The optimum condition for the strengthening of S–R connections is that the CR coincides with a reinforcing state of affairs. The greater the degree of temporal displacement of the reinforcing event and the CR (in either direction) the less effective the reinforcement.

The first statement modifies the strict contiguity principle maintained, for example, by Hull and by Guthrie by suggesting that a gradient operates both forward and backward in time. It is reasonable to suppose that the difference between connections based on strict contiguity and those based on lesser or greater degrees of discontiguity might differ quantitatively rather than qualitatively. The gradient can be used in this connection though the mechanism underlying it may not be specifiable at present. The question of whether the gradient of reinforcement is unidirectional or bidirectional is a controversial one. The problem is complicated by conflicting experimental evidence and by a confusion of empirical and theoretical statements. Neither of the statements made above is meant to be an empirical generalization; as assumptions they may be justified by their usefulness in prediction.

Historically, the delay of reinforcement principle belongs to instrumental learning where in some situations the experimenter can vary the period of delay by withholding the reward for any predetermined period of time after the subject makes the required response. He cannot do so in classical conditioning where, by definition, the UCS is presented in a fixed temporal relationship to the CS regardless of whether or when, the CR is evoked. The fact that the delay is not under direct experimental control does not, however, exclude the possibility that the delay of reinforcement principle applies in classical conditioning. Hilgard and Marquis (1940) recognize this and interpret classical conditioning in reinforcement terms but require that the CR be a preparatory reaction in which case the effect must be unidirectional.

The present interpretation is less specific than that of Hilgard and Marquis. It would seem best at present to follow Spence in defining a reinforcing state of affairs empirically as belonging to that class of "environmental events exhibiting the property of increasing the probability of occurrence of responses they accompany" (Spence, 1956, p. 32). If the reinforcing event is defined in this way, the CR does not have to be the direct means of bringing it about for strengthening to occur but may be facilitated if a reinforcing event occurs in that temporal neighborhood. Just what event in classical conditioning comprises a reinforcing state of affairs is a matter of discovery. However, because the UCS is the only event which is systematically associated with the S–R sequence it is reasonable to suppose that some feature of the UCS presentation is the reinforcer, most likely its onset or its offset. In conditioning fear reactions data presented by Mowrer and Solomon (1954) indicate that UCS onset is the critical event. Where the UCS is of constant (and short) duration either assumption will lead to essentially the same prediction and in the following discussion UCS onset is treated as the reinforcing event.

It is proposed to adopt tentatively the assumption of Hull (1952), Guthrie (1935), and of Hilgard and Marquis (1940) that the occurrence of a response at an effector is the event whose temporal relations to other events is crucial in establishing learned connections. While it is recognized that the latency of the UCR often differs from the latency of the CR and that both vary from trial to trial, for purposes of the present discussion the two will be treated as if equal and constant.

Consider the application of the theory to the limiting case, i.e., where the response latency is zero. If the CS–UCS interval is zero, the UCR starts at the onset of the CS and the CR starts with UCS onset. Therefore, simultaneous presentation of CS and UCS is most satisfactory as then both the gradient of contiguity and the gradient of reinforcement are at maximum strength. If the interstimulus interval is lengthened in either the forward or backward direction then both the relevant intervals are lengthened and conditioning is poorer.

In practice the response latency (x) is always greater than zero. In this case the CS–UCS interval which most favors contiguity between CS and UCR is $-x$ (that is, backward presentation where $\text{UCS} - \text{CS} = x$). However, if that is so, the CR is separated from UCS onset, the reinforcing state of affairs, by an interval of $2x$, so that reinforcement is relatively weak. Strongest reinforcement will occur where the CS–UCS interval equals $+x$, but then the CS and UCR are separated by an interval of $2x$ duration. So, if the interstimulus interval is such as to favor the effect of either the gradient of contiguity or of reinforcement, it necessarily follows that the other is operating at less than maximum efficiency. The longer the latency of the response the greater the relative disadvantage to the nonfavored mechanism.

In predicting the optimum interval the combination of overlapping gradients must therefore be considered. The predictions made will differ according to whether the manner of their combination is assumed to be additive or multiplicative. The multiplicative relationship assumes that both the gradients must be at greater than zero strength for conditioning to occur. That is, neither alone is sufficient for the establishing and/or strengthening of conditioned responses; both must operate in conjunction throughout the learning process. But the reinforcement principle is concerned with temporal relationships between the CR and the UCS. This gradient cannot, by definition, have an effect on an S–R tendency until the strength of the tendency is such as to bring the CR above threshold. That is, the effects of reinforcement are not felt until conditioning is already somewhat underway. How then, does the CS come to elicit the CR in the first place? The initial connection would appear to depend on the operation of the contiguity principle alone. Unless the arrangement of stimuli is such that the CS and UCR fall within the effective range of this gradient the CR will never be established. As the strength of the new connection reaches threshold value reinforcement begins to play a role. As the connection grows stronger and the response becomes more consistent the importance of reinforcement becomes correspondingly greater, until eventually the effect of contiguity is negligible by comparison. It is more meaningful, therefore, to adopt the assumption that the gradients combine in an additive fashion weighting the contribution of each in accordance with its changing importance.

The exact character of the predicted outcome depends on assumptions which are made concerning the shape, relative extent, and relative heights of the gradients. In the interests of simplicity and in the absence of direct evidence to the contrary it is assumed that both the contiguity and reinforcement gradients are linear. Though their actual extent is not specified it is envisaged that it would be measurable in terms of seconds or even milliseconds. For the present it is assumed, again for simplicity and in the absence of direct evidence, that both gradients are symmetrical and that they are equal in extent. The change in the relative importance of the two mechanisms is represented by changing the relative heights of the gradients according to the stage of training, while their extent remains constant throughout. Summarizing, the gradients are linear, symmetrical, and equal and limited in extent. The reinforcement gradient is at maximum height on the first trial (zero where there is no tendency for the CS to elicit the response to be conditioned before training begins) and rises as a function of the number of reinforced trials. This rise is accompanied by a fall in the height of the contiguity gradient indicating its relative decline in importance once the CR is established. The gradients combine in an additive fashion to affect the strength of the learned S–R connection, which is reflected in performance, other things being equal.

As the absolute extent of the gradients is not at present specifiable the amount of overlap for a response with a given latency is not known. However, relative statements can be made about the overlap for responses of short, medium, and long latencies. The overlapping gradients of contiguity and reinforcement are shown on the left of Figure 1 for responses of various latencies at various stages of training. The latency of the response is shown in the schematic representation by the amount of separation at the maxima of the gradients. Thus, Part A of Figure 1 represents the hypothetical case where latency equals zero. The upper dotted line represents the gradient of contiguity early and the gradient of reinforcement late in training. Conversely, the lower dotted line indicates the gradient of reinforcement early and that of contiguity late in training. The solid line represents two gradients when equal in height at an intermediate stage of training. Parts B, C, and D show the situation where the latencies are short, medium, and long respectively. If latency equals x then the distance between the maxima equals $2x$. Parts BI, BII, and BIII of Figure 1 show the change in relative contributions of the gradients at three stages of training, early, middle, and late, respectively. Thus, in Part BI, early in training, the height of the contiguity gradient exceeds that of the reinforcement gradient, in keeping with the argument that reinforcement is relatively ineffective at this stage.

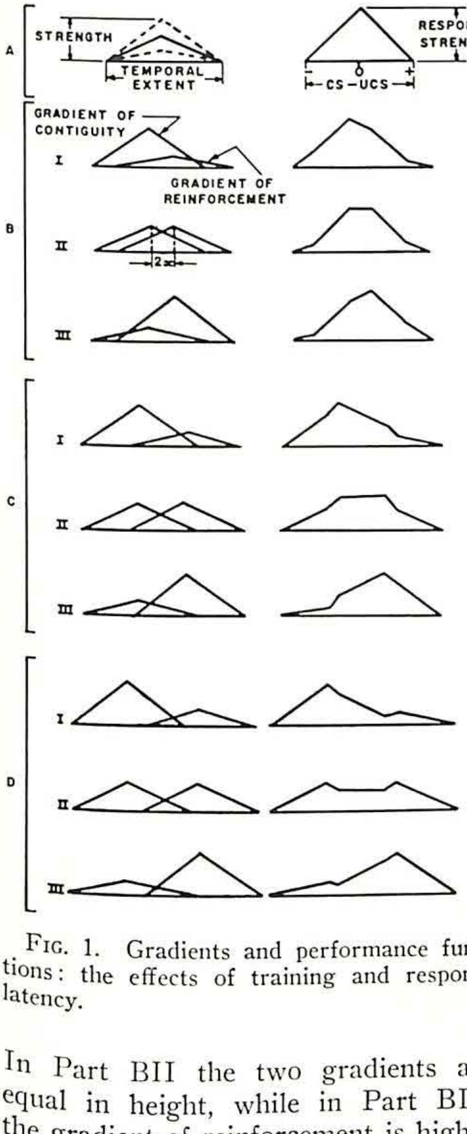


FIG. 1. Gradients and performance functions: the effects of training and response latency.

In Part BII the two gradients are equal in height, while in Part BIII the gradient of reinforcement is higher than the gradient of contiguity. Parts CI, CII, and CIII and DI, DII, and DIII represent the same changes in importance of the gradients for responses with longer latencies.

The corresponding performance functions relating response strength to CS–UCS interval are shown on the right of Figure 1, each of these functions being obtained by additive combination of the gradients on the left.[3]

Consideration of these performance functions leads to the following deductions from the theory:

1. With responses of latency greater than zero, the optimum interval for conditioning shifts with the amount of training (Parts BI, BII, and BIII). The direction of this shift is from a relatively short interstimulus interval early in training to a relatively long one later. The amount of shift is greater the longer the latency of the response being conditioned. (This shift in Parts BI–III may be compared with that in Parts CI–III and DI–III.)

2. With responses of such a latency that the overlap of the gradients is slight (Part D) a second peak appears in the performance function if the CS tends to elicit the response to be conditioned from the outset, as is the case where a weak response tendency is being strengthened rather than a new connection being formed. The longer the response latency, the greater is the separation of the peaks. Two experiments which fulfill these necessary conditions reveal the double-peaked effect. Wickens, Gehman, and Sullivan (1959) report 500 milliseconds best with a second peak at 1500 milliseconds and Jones (1961) obtained peaks at 440 and 1045 milliseconds. The conditions for producing the second peak are fairly common, in particular with "emergency" responses such as the GSR which was the response used in

---

[3] The value zero on the abscissae of the performance functions does not denote the simultaneous presentation of CS and UCS by the experimenter. The gradient of contiguity operates about a neural event consequent upon the occurrence of the CS rather than about the CS itself. In order to transform the functions shown here into forward and backward experimental arrangements of stimuli the zero point should be shifted to the left by an amount corresponding to the temporal lag between the CS and its neural counterpart. This distinction has not been made in labeling the figure as the predictions involve intervals of relative rather than of absolute lengths.

both the experiments cited. With many other responses, however, no such overt tendency is present before training and in some cases subjects who respond to the CS on a pretraining test are discarded. It might be argued that a covert tendency to respond is always present and all that is ever observed is the strengthening of existing response tendencies, rather than the establishing of new ones. However, it is assumed here that for strengthening by reinforcement to occur the response tendency must first rise above threshold so that an overt response, though perhaps a weak one, occurs in the temporal neighborhood of a reinforcing state of affairs. The mechanism which acts to bring the latent tendency above threshold is contiguity of the CS and the required overt response, this contingency being arranged experimentally by the presentation of the UCS. Where the CS and UCR are too widely separated the response tendency cannot be brought to the necessary strength in the experimental session. Hence, only when the tendency has already reached the requisite level, can the range of intervals spanned exclusively by the reinforcement gradient be effective in further strengthening the tendency in question. Hilgard and Marquis (1940) fail to indicate what mechanism acts to bring an initially subthreshold response tendency to sufficient strength for the operation of their reinforcement principle.

3. If measures of response strength are summed over all stages of training, the longer the response latency the wider the range of intervals which are effective in producing conditioned responses if the CS from the outset tends to elicit the response to be conditioned (compare the range of CS–UCS values embraced by the performance functions in Part B, C, and D). In addition, the performance function is less sharply peaked with longer latency responses. This prediction was tested by Jones (1961) in an experiment where two responses which were evoked by the same CS but differed in latency were conditioned in the same subject by the classical method. It was found that the function for the conditioned GSR was flatter and broader than that for finger movement.

4. Where the response latency is so long that the gradients do not overlap, the range of intervals covered by the contiguity gradient will give increased response strength corresponding to the height of that gradient early in training but the strength of the tendency will not continue to increase and may decline due to the lack of reinforcement. No other intervals will yield any conditioned responses unless the CS elicits the response to some extent from the outset. This case is not shown in Figure 1 but is an extension of Prediction 2 above.

The above predictions differ sharply from any to be derived from other approaches to the problem of the interval effect. Apart from the analysis by Hilgard and Marquis there has been no explicit attempt to relate the effect to response latency. It is possible, however, to deduce the appropriate interactions. Cognitive theory, which relates the efficacy of learning to the perceptual grouping of two stimulus events or the internal representations of such events, implies that characteristics of the responses evoked are irrelevant. Therefore, although it is difficult for a cognitive theorist to explain the superiority of a particular interval, he certainly expects to find the same interval superior whatever the latency of the CR or UCR. Hull is concerned with the contiguity of the stimulus trace and an effector activity as is clear from his statement of the primary reinforcement postulate (Hull, 1952, p. 5, Postulate III). Therefore, if the trace is at its maximum strength after 450 milliseconds, then: (*a*) an interstimulus interval of 450 milliseconds is superior only when the UCR has a latency approaching zero; (*b*) the longer the latency of the UCR the shorter should be the optimum CS–UCS interval; (*c*) where the response latency is longer than 450 milliseconds backward presentation of CS and UCS is most favorable; (*d*) optima longer than 450 milliseconds cannot occur regardless of the characteristics of the response. Guthrie also implies that a response rather than a neural event must be contiguous with the cue (which is almost certain to be a mediating stimulus event). It follows from Guthrie's analysis, as it does from Hull's, that the longer the latency of the UCR the shorter the optimum interval if the UCR is to be contiguous with a consistently occurring "true" CS. None of these positions are supported by the experimental evidence at present available.

The deductions from the two-principle theory may be formulated in terms of a comparison of trends in the learning curves for different interstimulus intervals. With response latency held constant, it is predicted that short-interval groups show a comparatively rapid initial improvement and the acquisition curves then flatten out and may even decline (see the rise and fall of the left wing of the performance function in Figure 1, Parts CI–III). Longer-interval groups (right-wing of the performance function) show a less rapid rise initially but eventually reach, and may even surpass, the shorter groups. The relationship between the curves is shown in Figure 2. The other theories mentioned above do not anticipate a change in pattern among the curves during acquisition but can predict only that the pattern established early becomes more distinct as training progresses. In a direct test of the prediction, Jones (1961) found the expected pattern among the acquisition curves for groups in which the GSR was conditioned at various intervals from 20 to 1245 milliseconds.

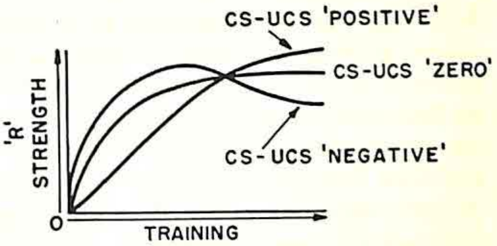


FIG. 2. Relationships among performance curves in the acquisition of a CR with various CS–UCS intervals.

Further evidence supporting the prediction is found in the learning curves obtained by Moeller (1954) for the GSR and McAllister (1953) for the eyeblink which show that longer-interval groups approach their performance asymptote at an initially slower rate than shorter-interval groups. Jones (1959) observed that groups in which the GSR was conditioned with 550 and 1060 milliseconds intervals respectively approached approximately the same asymptote but that the 550 milliseconds group did so much more rapidly. It may be noted that White and Schlosberg (1952), who report a short optimum of 500 milliseconds for the GSR using measures taken on extinction trials, trained their subjects for but five conditioning trials.

It is often reported of simultaneous and backward presentation of the CS and UCS that the performance curves show a brief initial rise followed by a gradual decline (Spooner & Kellogg, 1947). This is the type of effect that would be expected under these conditions from the analysis given here.

The explanation commonly offered invokes a new mechanism sometimes referred to as "sensitization" but no additional concept is required to incorporate the effect here. Backward conditioning emerges as a special case of the interval effect rather than differing qualitatively from forward conditioning. Hull (1943, p. 172) suggested a treatment of this kind in his discussion of Wolfle's data but his later development of the stimulus trace concept made such interpretation difficult where the UCR as well as the UCS preceded the CS. Confirming the possibility of real backward conditioning, Champion and Jones (1961) found that a group in which the GSR was conditioned with a backward interval of 750 milliseconds performed better than did a pseudoconditioning control group.

Another prediction which follows from the present interpretation concerns the effect of a shift of interstimulus interval during the course of training. Suppose that for a group of subjects forward conditioning begins with a short interval in the range most favorable for the effect of contiguity until the CR begins to appear fairly consistently. If at this stage the interval is changed to a longer one, more favorable for the effect of reinforcement, performance will surpass that of control groups conditioned throughout at either the shorter or the longer interval. Further, long latency responses are able to tolerate larger shifts of this kind than short latency responses.

If the CS–UCS interval is held constant and the learning curves for responses of different latencies are compared they will differ in shape in a manner depending on the range of values in which the interval is held. At an intermediate range of interval values the curves for different responses have the same shape because contiguity and reinforcement are favored equally regardless of response latency. However, with intervals in the shorter range (e.g., as in backward conditioning) separation of CR and UCS onset (reinforcement) is much greater than the separation of CS and UCR (contiguity) and the difference between the two is emphasized with a long latency response. As it has been proposed that contiguity is of greatest importance early in training but that reinforcement gains in importance during the later stages, the longer latency response shows the more rapid initial rise, but later in training the shorter latency response reaches and maintains a higher level of performance. Conversely, if the interval is lengthened beyond the intermediate range (as in forward conditioning) then contiguity is more adversely affected than is reinforcement. This difference is again emphasized where the response has a long latency. Therefore, in this case the performance curve for the short latency response shows a more rapid initial rise but is later surpassed by that for the long latency response. The predicted relationship was obtained by Jones (1961) in conditioning finger withdrawal and GSR with forward intervals.

The two-principle theory of classical conditioning put forward in this paper leads to the set of predictions outlined above, some of which have already been confirmed. These predictions cannot be derived from other theories in their present form. No theorist to date has discussed the interval effect in relation to the course of learning and only Hilgard and Marquis have attempted to relate the optimum CS–UCS interval to the latency of the response being conditioned. When these factors, i.e., amount of training and response latency, are considered in terms of the constructs offered by others in explanation of the interval effect the predictions of the interactions concerned differ markedly from those of the present interpretation.

SUMMARY

The function relating the interval between the onset of the CS and the UCS to response strength in classical aversive conditioning is interpreted in terms of the joint action of two principles. Early in training the efficacy of conditioning depends mainly on the contiguity of CS and UCR but later mainly on the proximity of the CR to reinforcement (UCS). In both cases a bidirectional gradient is assumed to operate such that the greater the discontiguity of the events the less effective the learning. It is predicted that variations in experimental procedure, such as the latency of the response being conditioned, which affect either one or both of these relationships affect the function relating the CS–UCS interval to response strength and that the shape of the learning curves depends on the CS–UCS interval used in training.

REFERENCES

BERNSTEIN, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.

BIERBAUM, W. B. The temporal gradient in GSR conditioning. *J. gen. Psychol.*, 1958, **59**, 97–103.

CHAMPION, R. A., & JONES, JOAN E. Forward, backward, and pseudoconditioning of the GSR. *J. exp. Psychol.*, 1961, **62**, 58–61.

GERALL, A. A., & WOODWARD, J. K. Conditioning of the human pupillary dilation response as a function of CS-UCS interval. *J. exp. Psychol.*, 1958, **55**, 501–507.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.

HANSCHE, W. J., & GRANT, D. A. Onset versus termination of a stimulus as the CS in eyelid conditioning. *J. exp. Psychol.*, 1960, **59**, 19–26.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century 1940.

HULL, C. L. *Principles of behavior: An introduction behavior theory.* New York: Appleton-Century, 1943.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

JONES, JOAN E. CS-UCS intervals in GSR conditioning. Unpublished PhD thesis, University of Sydney, 1959.

JONES, JOAN E. The CS-UCS interval in conditioning short and long latency responses. *J. exp. Psychol.*, 1961, **62**, 612–617.

KIMBLE, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1–15.

MCALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–423.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

MOWRER, O. H. On the dual nature of learning: A re-interpretation of "conditioning" and "problem-solving." *Harvard educ. Rev.*, 1947, **17**, 102–148.

MOWRER, O. H., & SOLOMON, L. N. Contiguity vs. drive-reduction in conditioned fear: The proximity and abruptness of drive-reduction. *Amer. J. Psychol.*, 1954, **67**, 15–25.

MYERS, J. An experimental investigation of the effect of varying the time between the onsets of the conditioned and unconditioned stimuli on the conditioned eyelid response. Unpublished PhD thesis, State University of Iowa, 1950.

REYNOLDS, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35**, 15–30.

SCHLOSBERG, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44**, 379–394.

SKINNER, B. F. Two types of conditioned reflex and a pseudo-type. *J. gen. Psychol.*, 1935, **12**, 66–77.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPOONER, A., & KELLOGG, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **60**, 321–334.

TOLMAN, E. C. *Purposive behavior in animals and men.* Berkeley: Univer. California Press, 1949.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of period of delay. *J. exp. Psychol.,* 1952, **43**, 357–362.

WICKENS, D. D., GEHMAN, R. S., & SULLIVAN, S. N. The effect of differential onset time on the conditioned response strength to the elements of a stimulus complex. *J. exp. Psychol.,* 1959, **58**, 85–93.

WOLFLE, H. M. Time factors in conditioning finger withdrawal. *J. gen. Psychol.,* 1930, **4**, 372–378.

WOLFLE, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.,* 1932, **7**, 80–103.

(Received September 12, 1960)

Psychological Review
1962, Vol. 69, No. 3, 187–201

PATTERN AND LEVEL OF RISK IN GAMBLING DECISIONS[1]

DEAN G. PRUITT
University of Delaware

This paper introduces a new deterministic model for analyzing decision making in gambling, the pattern and level of risk (*PLR*) model. At the beginning of the paper, the experimental evidence for four traditional models is examined and the inadequacies of these models shown. Then the *PLR* model is described, predictions are drawn from the model, and evidence is presented to support the predictions. Finally some implications of the *PLR* model are shown.

THE EVIDENCE FOR TRADITIONAL MODELS

Edwards (1955) described four, related deterministic models which have, in large part, guided experimentation on gambling decisions. All four assume that a gambler will choose the alternative (e.g., a bet) with the maximum value, and that the value of an alternative can be calculated by multiplying the value of each of its outcomes by the probability that the outcome will occur and summing these products over all outcomes. The models differ with respect to whether objective or subjective values and probabilities are used in these calculations. The four models are defined by the following equations for calculating the value of an alternative:

$$EV = \sum_i p_i \$_i \qquad [1]$$

$$SEM = \sum_i P_i^* \$_i \qquad [2]$$

$$EU = \sum_i p_i u_i \qquad [3]$$

$$SEU = \sum_i P_i^* u_i \qquad [4]$$

Equation 1 defines expected value (EV), where p_i and $\$_i$ stand for the objective probability and dollar value of the *i*th outcome and Σ/i means that the products $p_i\$_i$ are summed over all of the potential outcomes of the alternative. Equation 2 defines subjectively expected money (SEM), where P_i^* refers to the subjective probability of the *i*th outcome and the other symbols have the same meanings as above. Equation 3 defines expected utility (EU) where u_i refers to the utility (subjective value) of the *i*th outcome and the other symbols have the same meanings as above. Equation 4 defines subjectively expected utility (SEU); all of the symbols have the same meanings as in the other models.

In evaluating these models, three criteria will be kept in mind: (*a*) the range of alternatives or choice situations within which each model has predictive value; (*b*) the accuracy with which each model predicts decisions within its range; and (*c*) in ranges where both have predictive value, whether a model containing a subjective parameter is any better than the comparable one containing an objective parameter. The last considera-

---

[1] This paper was completed during tenure of a United States Public Health Service postdoctoral research fellowship as a Research Associate in the Program of Graduate Training and Research in International Relations at Northwestern University.

The author is indebted to C. H. Coombs, Ward Edwards, Harold Guetzkow, and John D. Pruitt for their advice.

tion is important from the point of view of parsimony. If a model with an objective parameter is as good at predicting decisions within a certain range as the comparable model with a subjective parameter, then the former should be preferred because it requires no special measurement procedures.

The EV *Model*

The *EV* model (and the *SEM* and *EU* models as well) is, of course, inapplicable in situations where the objective parameters which compose it are not defined. In addition, it is unlikely to be valuable where objective parameters, though defined, are not in some way communicated to the gambler or where individual tastes are obviously important in determining the value of outcomes (as in preference among foods). There remains a hard core of situations in which objective probabilities are fairly apparent and outcomes are money or objects easily convertible into money. Here the *EV* model is at least worthy of consideration. Even when these conditions are met, the *EV* model is known to be grossly inadequate in the range of bets which have very small probabilities of winning or losing large amounts of money (Allais, 1953; Bernoulli, 1954). The inaccuracy of the *EV* model in these cases is shown by the willingness of people who are well informed about the probabilities to buy lottery tickets and insurance (where *EV* is considerably less than zero) and to play the St. Petersburg Paradox only at very low stakes (where *EV* is infinite at any stake).

In the range of bets which have moderate probabilities and values of outcomes, research findings seem to support the following generalization: The *EV* model is a poor predictor of decisions where *EV* differences are small, but improves in predictive capacity as *EV* differences increase. Thus, the model is useless for predicting choices between bets with the same *EV;* the model predicts indifference between such bets whereas many experiments have revealed consistent preference (e.g., Coombs & Komorita, 1958; Edwards, 1953). Mosteller and Nogee (1951) found that prediction improved with an increase in *EV* difference as follows: Where the difference in *EV* between two two-outcome bets lay between 2¢ and 50¢, only 42% of the predictions were correct (50% would be expected by chance); between 55¢ and $1.55, 57% were correct; between $1.80 and $2.20, 60% were correct; and between $2.50 and an undisclosed maximum, 67% were correct. These percentages would undoubtedly have improved somewhat if the bets had been replicated and "dominant" preferences established; but they are low enough to suggest that the *EV* model leaves much to be desired as a predictive device.

One other generalization about the scope of the *EV* model is suggested by an experiment by Edwards (1954a): The less complex the bets between which a decision is made, the more accurate will the *EV* model be in predicting these decisions.

The SEM *Model*

There is one kind of bet, very common in everyday gambling situations such as roulette or horseracing, in which the *SEM* model makes grossly incorrect predictions. This is the bet in which a person may stake as much money as he likes on a chance event and wins a fixed multiple of his stake if the chance event occurs. In such cases, the *SEM* model (and the *EV* model) predicts that if a man will bet at all on a chance event, he will put all the money he has on that chance event, a prediction which is incompatible with everyday observations of gamblers. However, in the typical laboratory situation, where only two bets are available at a time, the model does not lead to absurd predictions.

Laboratory data appear to support two generalizations about the *SEM* model: (*a*) The *SEM* model is fairly accurate in predicting choice between simple bets which differ in probability of winning or losing but whose outcomes involve the same general level of money. (*b*) The *SEM* model is only slightly if any better than the *EV* model in its ability to predict choice between bets with the same probabilities of winning and losing but differing in level of money.

Evidence for the first generalization comes from the work of Edwards on probability preferences (1953, 1954a, 1954b, 1954c). When asked to choose between bets with equal *EV* but differing in probability of winning, the subjects showed a marked group preference for certain probabilities (e.g., 4/8) and a marked dislike for others (e.g., 6/8). The same preference patterns were found when the amount of money involved in the bets was increased, making an interpretation in terms of the *EU* model very unlikely. Similar results were found for individual cases by Coombs and Pruitt (1960), who pointed out in addition that people differ in the probabilities they prefer. The *SEM* model would appear to be the simplest way of explaining these preferences, and a test of this explanation was made by Edwards (1955). First he used a betting method which assumed the *SEM* model to measure for each subject the subjective probability of winning and losing at eight levels of probability. Then he inserted these values into the *SEM* model to predict probability preferences. Eighty-three percent of the choices between winning bets and 65% of the choices between losing bets were successfully predicted. Other evidence suggests that predictions would have been even better if the choices had been replicated enough times to establish the dominant probability preference pattern for each subject. It should be noted that the *EV* model is clearly inadequate to explain these data since all of the bets had the same *EV* level so that there was no basis for making predictions.

Evidence for the second generalization comes from an experiment by Suppes and Walsh (1959) in which predictions by the *SEM* model to choices between two-outcome bets with a subjective probability of winning of $\frac{1}{2}$ were correct only 57% of the time.

The EU *Model*

The measurement of utility has turned out to be a very difficult and elusive task. At least seven methods have been used over the past 10 years, three assuming the *EU* model and four the *SEU* model. In each case the utility values thus measured have been put back into the model to predict choices between pairs of bets. Thus, to a greater or lesser extent, these studies constitute a test of the *EU* and *SEU* models.

Two of the three *EU* experiments have not been adequate tests of the model. One, by Edwards (1955) used a method of measuring utility which he has since repudiated (1961); the results were inconclusive. The other by Coombs and Komorita (1958) yielded 29 out of 30 correct predictions, but all 30 would have been correct on the assumption that the subjects preferred less risk to more, a characteristic they showed throughout their behavior. Furthermore only three subjects were used.

The third study (Mosteller &

Nogee, 1951) is a better test of the model but there are ambiguities in interpreting it. (Some of the results of this study have already been presented in the section on the *EV* model.) Utilities of money were generated and used to predict choices between bets at a number of different levels of probability and money. These predictions were correct in 66% of the pairs, as opposed to 50% success with the *EV* model (expectation by chance was also 50%). As in the case of predictions from the *EV* model, the accuracy of prediction increased with an increase in *EU* difference between the bets: in pairs where the difference was over \$2.50, 93% of the predictions were correct. One might conclude from this study that: The *EU* model is somewhat better than the *EV* model when the *EU* difference between bets is small and improves markedly as the *EU* difference increases. However, the method used for measuring utility is ambiguous; the decisions made by the subjects in the utility measuring sessions could just as well have been determined by subjective distortions of probability as by departure of the utilities from money values. If this had been true, the *SEM* model would have made as good or better predictions than the *EU* model; this possibility was not examined, although data were available to do so. Conclusions about the value of the *EU* model must await this further check.

The SEU *Model*

Aside from Edwards' (1955) experiment mentioned in the previous section, which employed a questionable method for measuring utility and will therefore not be further discussed, all the experiments on the *SEU* model have used a chance event with both objective and subjective probability equal to $\frac{1}{2}$. This means two things: (*a*) that the model remains untested over a broad range of bets and (*b*) that the *EU* and *SEU* models cannot be compared on the basis of existing data.

In the case of bets involving money, the *SEU* model does not appear to have been particularly successful. One study (Davidson, Suppes, & Siegel, 1957, pp. 19–81) claims success, but the test of the *SEU* model was very weak permitting quite variable results to be considered positive evidence; the results must be considered inconclusive. The other experiment, done by Suppes and Walsh (1959), was a much stronger test of the model. The *SEU* model predicted correctly only 58% of the choices between paired comparisons (chance expectation: 50%) and was only slightly better than the *SEM* model which predicted correctly 57% of the choices.

The *SEU* model made a somewhat better showing in one study which used bets composed of outcomes other than money. Davidson, Suppes, and Siegel (1957, pp. 82–103) measured the utility of phonograph records with a linear programming method. Depending on certain minor assumptions, predictions about choices between bets containing these phonograph records were correct 67% or 71% of the time. Another study (Hurst & Siegel, 1956) used cigarettes as the outcome, but the number of correct predictions was not reported.

Summary Evaluation of the Traditional Models

All four of the traditional models have shown some success in predicting choices between bets, but most of the successes have been of moderate magnitude (57% to 71% correct). There are two possible exceptions to this generalization, each involving restricted ranges of bets: (*a*) When positive expected value bets differ in probability of winning but have roughly the same level of money in their outcome, the *SEM* model is quite successful (83% correct). (*b*) When *EU* differences are very large, the *EU* model seems to do a very good job (93% correct) but there are some ambiguities in interpreting these findings. Aside from these limited findings, the four models have shown only moderate success.

WHERE DO WE GO NOW?

There are two directions which the study of gambling behavior might take from here: (*a*) further refinement of the traditional models, e.g., new attempts to measure utility and subjective probability, more comprehensive studies of the range in which these models have predictive value, or (*b*) search for new models. The first direction has much to be said for it; there are many gaps in our knowledge; most of the generalizations stated above could stand re-evaluation with new data. But there are other considerations which strongly suggest the appropriateness of the second direction: In an infant science, such as the study of decision making, most approaches are likely to be wrong so that the need for innovation is likely to be greater than the need for model refinement. Furthermore the results discussed above are not really encouraging—a great deal of the variance is still not explained—and there is no reason to believe that the traditional models will be able to account for it. In view of these arguments, the author has opted for a search for new models.

This does not mean that we should utterly reject the traditional models; at this point they are the best we have. Because they are the best we have, they can and must be used in building models of other situations which resemble gambling (for examples see Siegel, unpublished[2]; Snyder, 1960; Tanner & Swets, 1954). But some effort must also be made to improve our understanding of gambling.

The search for new directions, which has culminated in the *PLR* model of this paper, was begun by Coombs and the author (1960) in a study of the utility of risk. In this study, utility of risk was conceptualized in terms of "variance preference": the greater a person's utility of risk the higher variance he will prefer. A partial model of variance preference was devised: that the order of preference among variances for any given probability of winning will be a folded **J** scale of variance. And this partial model was overwhelmingly supported by the data. This and some other regularities of behavior found in that study plus some observations of real life gambling have led inductively to a more comprehensive partial model, the *PLR* model.

THE PATTERN AND LEVEL OF RISK MODEL

The model which is developed in this paper applies only to bets which have a least one negative outcome, i.e., which have some element of risk. In the discussion section, some speculations will be presented about the possibility of generalizing it to all bets.

Concepts

Every bet which has at least one negative outcome can be completely described by specifying its "pattern of risk" and "level of risk." The pattern of risk ("pattern" for short) of a bet is determined by the number of possible outcomes, the probability of achieving each outcome and the ratio of the outcomes to one another. Two bets have the same pattern if they have the same number of outcomes at the same probabilities and if the outcomes of one can be obtained by multiplying the corresponding outcomes of the other by a constant. Thus two bets with the same pattern are:

---

[2] Siegel, S. Decision making and learning under varying conditions of reinforcement. Unpublished manuscript, 1960.

A: $\frac{1}{2}$ chance to win 50¢
$\frac{1}{2}$ chance to lose 50¢

B: $\frac{1}{2}$ chance to win \$1
$\frac{1}{2}$ chance to lose \$1

Two others are:

C: 1/2 chance to win \$1
1/3 chance to lose 75¢
1/6 chance to lose 60¢

D: 1/2 chance to win \$3
1/3 chance to lose \$2.25
1/6 chance to lose \$1.80

The level of risk of a bet is defined as the sum of its negative outcomes weighted by their respective probabilities of occurrence. Thus the level of risk of Bet A is 25¢; of Bet B, 50¢; of Bet C, 35¢; and of Bet D, \$1.05.

The concepts of pattern of risk and level of risk may seem arbitrary at first: A bet can be described in many terms; why choose these? It is the contention of the author that people usually perceive bets in these terms. The evidence for this lies in the structure of most popular gambling situations, where characteristically there is a sharp distinction between the chance events and payoff ratios on the one hand (pattern of risk) and the size of stake on the other (level of risk). For instance, in roulette each of the squares on which stakes may be placed corresponds to a specific chance event and has a fixed ratio of payoff to stake; thus each of the squares corresponds to a specific pattern of risk. The size of the gamblers stake, which determines the level of risk, is completely distinct from the square on which he puts it. In betting on horse races, the gambler chooses a pattern by deciding which horse to bet on and a level of risk by deciding what size ticket to buy. In friendly wagers, separate announcements are often heard of the pattern of the wager (e.g., "I'll give you 5 to 3 on Navy") and the amount bet ("O.K., I'll put \$3 on Army."). Finally, in investment a sharp distinction is drawn between the company or bond invested in (pattern) and the number of shares or bonds bought (level of risk).

In building the *PLR* model, it is assumed that people have separate attitudes toward pattern and level of risk and that these interact to determine gambling decisions. Attitude toward pattern is embodied in the concept "utility of pattern," each pattern being assumed to have its place on a ratio scale of utility. In cases where utility of pattern is measured on an ordinal scale, we may speak in terms of pattern preferences (e.g., "Pattern X is preferred to Pattern Y."). Attitude toward level of risk is embodied in the concept "utility of risk," which is assumed to be measurable on a ratio scale.

The PLR *model*

At this stage in the development of the pattern and level of risk model (*PLR* model), no assumptions have been made about the nature of the relationship between utility of pattern and pattern itself other than the assumption that this relationship differs from person to person. Some speculations about this question will be presented in the discussion section. The function relating utility of risk to level of risk is also assumed to differ from person to person but to have in common for all people three characteristics: (*a*) being negative for all levels of risk, (*b*) being a negatively accelerated decay curve, and (*c*) being zero when the level of risk is zero.

The *PLR* model is analagous to the traditional models in as much as it provides a formula for computing the utility of alternatives and assumes that the gambler will choose the alternative which has the greatest utility. This formula may be stated as follows:

$$U(X) = r_X \cdot u(p_X) + g(r_X) \quad [5]$$

where:

$U(X)$ is the utility of alternative X.
r_X is the level of risk involved in X.
$u(p_X)$ is the utility of the pattern p_X of X.
$g(r_X)$ is the utility of risk.

Causality moves in only one direction in this model: $U(X)$ is the dependent variable and the terms on the right of the equation are independent variables. Again it should be emphasized that the *PLR* model applies only to bets which have at least one negative outcome.

The model is illustrated in Figure 1, where $U(X)$ is shown as a function of r_X. An intuitive account of the model follows. In order to discuss the model intuitively, a name will be needed for the first term on the right hand side of Equation 5. This is called, heuristically "expectation of gain"; the greater this term, the more the gambler "expects to get out of a bet." The term has the following properties, all intuitively compelling: As the utility of a pattern $u(p_X)$ increases, expectation of gain from bets embodying that pattern also increases. In the case of patterns whose utility $u(p_X)$ is positive, expectation of gain will increase as the amount of money risked r_X increases, since for a given pattern, increasing the amount risked also increases the amount which may be won. But where the utility of a pattern is negative, expectation of gain decreases as the amount of money risked increases.

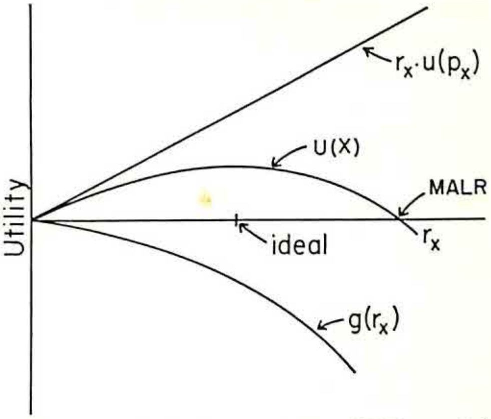


FIG. 1. Illustration of the *PLR* model showing the utility of an alternative $U(X)$ as a function of the level of risk involved in that alternative r_X. Also shown are the components of $U(X)$, the expected gain $r_X \cdot u(p_X)$ and the utility of risk $g(r_X)$. (The ideal level of risk, which comes at the point where $U(X)$ is maximal, and the maximum acceptable level of risk, MALR, are indicated.)

Expectation of gain and utility of risk are conceived intuitively as hypothetical constructs, having genuine referents in the form of two conflicting or cooperating behavior tendencies within the individual. These tendencies cooperate when the utility of a pattern is negative. When the utility of pattern is positive, however, these tendencies come in conflict and the conflict increases as the level of risk increases. Because the utility of risk is a negatively accelerated decay curve over level of risk, the model predicts what we know from experience (and what the *SEM* model failed to predict), that there is a limit to the amount of money a man will risk on a pattern which has at least one negative outcome, no matter how favorable this pattern may appear.

Derivations from the Model

In this section will be presented five propositions drawn from the *PLR* model. All of them refer to the behavior of individuals rather than groups.

Proposition 1. *At any level of risk, the order of preference among patterns of risk will be the same as at any other level of risk.*

Proof. Take any two bets X and Y at level of risk R, where X has pattern p_X and Y has pattern p_X and assume that X is preferred to Y, i.e., $U(X) > U(Y)$. By Formula 5, this implies that:

$$R \cdot u(p_X) + g(R) > R \cdot u(p_Y) + g(R)$$

which simplifies to:

$$u(p_X) > u(p_Y)$$

Now take any other level of risk S and find the two bets X' and Y' which have the same patterns, p_X and p_Y. Since we know that:

$$u(p_X) > u(p_Y)$$

it follows that:

$$S \cdot u(p_X) + g(S) > S \cdot u(p_Y) + g(S),$$

which implies by Formula 5 that $U(X') > U(Y')$, i.e., that X' is preferred to Y'. Thus it is shown that the order of preference between any two patterns of risk at any level of risk will be the same as the order of preference between the same patterns at any other level, from which Proposition 1 follows.

Note. In the next four propositions, two new terms are introduced: the "ideal level of risk," which is the level of risk which the individual most prefers for a given pattern of risk, and the "maximum acceptable level of risk," which is the highest level of risk he will voluntarily accept for a given pattern.

Proposition 2. *The more preferred a pattern of risk, the higher will be the ideal level of risk for that pattern.*

Proof. This proposition will be demonstrated with reference to Figure 1. The ideal level of risk is found, of course, at the point where the curve $U(X)$ reaches a maximum. It should be immediately apparent that a person can have only one ideal level of risk since the sum of a straight line and a curve which is negatively accelerated is itself a negatively accelerated curve and, therefore, can have at most one maximum. By calculus, this maximum occurs at the level of risk where:

$$U'(X) = u(p_X) + g'(r_X) = 0$$

i.e., where:

$$u(p_X) = -g'(r_X)$$

To put this in other words, the ideal level of risk for a pattern is found where the utility of that pattern equals the negative slope of the utility of risk curve.

Now assume two patterns of risk P and Q such that $u(P) > u(Q)$, i.e., that P is more preferred than Q. According to our calculus, the negative slope of the utility of risk curve will be greater at the ideal for P than at the ideal for Q. Since the negative slope of the risk curve increases with increasing level of risk, this means that the ideal for P will be found at a higher level of risk than the ideal for Q.

Proposition 3. *The more preferred a pattern of risk, the higher will be the maximum acceptable level of risk for the pattern.*

Proof. This proposition will also be demonstrated with reference to Figure 1. It is well known that a curve which is negatively accelerated throughout can be crossed by a straight line at most two times. In this case, the function $U(X)$ of r_X is the negatively accelerated curve and the level of risk axis (the abscissa) is the straight line. By definition these curves cross one another at the origin. The other crossing, if it exists, is the maximum acceptable level of risk, i.e., the point where $U(X) = 0$. By algebra, the maximum acceptable level of risk for a pattern p_X occurs at the point where:

$$u(p_X) = -g(r_X)/r_X$$

i.e., where the utility of this pattern equals the negative utility of risk divided by the level of risk.

Now assume two patterns of risk P and Q such that $u(P) > u(Q)$, i.e., that P is preferred to Q. According to our formula, this means that the ratio of the negative utility of risk divided by the level of risk will be greater at the maximum for P than at the maximum for Q. Since the function relating the utility of risk to level of risk is a negatively accelerated growth curve, this ratio will be greater the greater is the level of risk. Therefore the maximum level of risk for P will be higher than the maximum level of risk for Q.

Proposition 4. *For any given pattern of risk, divide the levels of risk into two groups: those smaller than the individual's ideal and those greater than his ideal. Within each group, levels of risk will be more preferred the closer they are to the ideal level of risk.*

Proof. As was mentioned above under Proposition 2, the ideal level of risk occurs at the point where $U(X)$ reaches a maximum. Since $U(X)$ is negatively accelerated throughout, it can have at most one maximum and no minima. The proof of Proposition 4 comes from the nature of such a maximum . . . namely, that the curve rises ever higher as it approaches the maximum and declines ever more sharply as it proceeds beyond the maximum. This implies a theorem that *on either side of the ideal the utility of bets with a given pattern increases as the level of risk approaches the ideal level of risk.* Proposition 4 is a restatement of this theorem in preference language.

Proposition 5. *For any given pattern of risk, all levels of risk up to the maximum acceptable level of risk will be acceptable.*

Proof. This proposition follows from the assumption, stated in the discussion of Proposition 3, that the function of $U(X)$ of r_X can cross the level of risk axis at most two times, the first of which is the origin and the second, if it exists, the maximum acceptable level of risk. Since it cannot cross the level of risk axis between these points, it remains positive between them . . . all levels of risk up to the maximum level have a positive $U(X)$, which means that they are all acceptable.

Experimental Evidence for the Propositions

The experimental evidence presented in this section is derived from two experiments on gambling.

The method and part of the results of the first experiment are reported more extensively elsewhere (Coombs & Pruitt, 1960). The procedure consisted of presenting a large number of forced choices between two-outcome bets, such as the following:

$$\text{Bet A: } \begin{matrix} 1/3 \text{ to win } \$1.40 \\ 2/3 \text{ to lose } 70 \text{ cents} \end{matrix}$$

$$\text{Bet B: } \begin{matrix} 1/2 \text{ to win } \$1.00 \\ 1/2 \text{ to lose } \$1.00 \end{matrix}$$

All bets had an expected value (EV) of zero. The bets were organized into five sets. Those in Sets I, II, and III had a constant probability of winning and differed in variance.[3] The probability of winning for Set I was 1/3; for Set II, 1/2; and for Set III, 2/3. There were six levels of variance in each set: .12, .30, 1.0, 5.0, 25.0, and 100.0. The bets of Sets IV and V had a constant variance and differed in probability of winning. Set IV had a variance of 1.0 and Set V a variance of 25.0. The same levels of probability were used in both sets: 1/6, 1/3, 1/2, 2/3, and 5/6. All of the bets within each set were paired with one another and each of the resulting pairs was presented eight times. Ninety-nine undergraduates were used as subjects.

---

[3] The variance of bets of this type is computed by the formula $\sigma^2 = pq(a-b)^2$, where p and q are the probabilities of winning and losing respectively, and a and b are the amounts to be won and lost.

The second experiment was similar to the first in most respects but had six sets. Again the bets in Sets A, B, and C had a constant probability of winning and differed in variance. Set A had a probability of 1/3, Set B of 1/2, and Set C of 2/3. Seven levels of variance were employed in each set: .12, 2.0, 5.1, 12.5, 21.2, 32.0, and 45.2. The bets in Sets D, E, and F had a constant variance and differed in the probability of winning. Set D had a variance of .12, Set E of 12.5, and Set F of 45.2. In each set there were three levels of probability: 1/3, 1/2, and 2/3. As in the case of the first experiment, the bets were exhaustively paired within Sets D, E, and F, and each pair was presented eight times. However, in the case of Sets A, B, and C, an accept-reject method was used instead of paired comparisons; each bet was presented eight times and the subject was asked in each case to decide whether or not to accept the bet. Thirty-nine undergraduates were used as subjects.

Before discussing the results, three things must be said about the relationship between these experimental operations and the terms of the *PLR* model. (*a*) It can be easily shown that for two-outcome bets with *EV* of zero, the pattern of risk is completely determined by the probability of winning. Thus, the five probabilities of winning used in these experiments correspond to five patterns of risk: 1/6, 1/3, 1/2, 2/3, and 5/6. (*b*) Within the range of probabilities used in these experiments, two bets with different probabilities of winning (different patterns, that is) and the same variance will have roughly the same level of risk. And level of risk is monotonicly related to variance. Thus, the six levels of variance in the first experiment correspond roughly to six levels of risk, and the seven levels of variance in the second experiment to seven levels of risk. (*c*) the design of these experiments permits us to make three kinds of analysis: In the case of stimulus Sets I, II, III, A, B, and C, we can study preferences among levels of risk while holding the pattern of risk constant. In the case of Sets IV, V, D, E, and F, we can study pattern preferences while holding level of risk constant. Finally, for each subject, we can compare his preferences among levels of risk with his pattern preferences to discover interactions.

Proposition 1. At any level of risk, the order of preference among patterns of risk will be the same as at any other level of risk.

The responses to Sets IV and V in the first experiment provide us an opportunity to compare each subject's order of preference among the five patterns (1/6 to 5/6) at a low level of risk (variance of 1.0) with his order of preference among the same five patterns at a high level of risk (variance of 25.0). Proposition 1 implies that these two preference orderings will be identical.

Preference orderings were obtained from the data by assuming that subject preferred Bet A to Bet B if he chose A on more than half of the occasions when it was paired with B. Inspection of the results revealed striking similarity between the orders of preference at the two levels of risk, as predicted; but it is necessary to express this similarity in a quantitative fashion. An index was therefore devised for expressing these preference orderings, which ranged from a low of 1, for a pattern of consistent preference in all pairs for the smaller probability of winning, to a high of 11, for a pattern of consistent preference for the larger probability of winning.[4] Two such ratings were computed for each subject, one to describe his pattern preferences in Set IV, the other to describe his preferences in Set V. The correlation, over subjects, between these two ratings was .93, a highly significant result, indicating considerable stability in the order of preference among patterns as one moves from a low to a high level of risk.

Similar results were found in the second experiment for Sets D, E, and F.

Proposition 2. The more preferred a pattern of risk, the higher will be the ideal level of risk for that pattern.

Since all of the levels of risk within a set were exhaustively paired, it was possible to identify for each of the Sets, I, II, and III of the first experiment that level of risk which was preferred over all the others, i.e., the ideal level of risk. This meant than an ideal level of risk was identified for three patterns: 1/3 (Set I), 1/2 (Set II), and 2/3 (Set III). An index was devised for expressing these ideals, which ranged from 1, indicating a preference for the lowest level of risk, to 16, indicating a preference for the highest available risk. (Since there exist levels of variance below .12 and above 100.0, this scale has an arbitrary ceiling and floor.) Preferences between the three patterns were ascertained from the data of Sets IV and V combined.

Proposition 2 was tested in the following form: If one pattern is preferred to another, there will be a higher ideal level of risk for the former than for the latter. Data supporting this hypothesis are shown in Table 1.

In Table 1a the subjects are divided into two groups (columns) on the basis of which pattern they preferred, 1/3 or 1/2. (Two subjects are omitted who did not have a preference between the two patterns.) Each of these two groups is divided into three subgroups (rows) depending on whether the ideal level of risk for Pattern 1/3 was greater than, equal to, or less than the ideal for Pattern 1/2. The results show a strong tendency for the more preferred pattern to have a higher ideal level of risk, as hypothesized. $\chi^2 = 23.13$ for the contingency table, with $p < .01$. The relatively large number of cases in which both ideals were equal might be thought to weaken this hypothesis. However, for most of them (25 out of 30), the most preferred level of risk for both patterns was at one or the other of the extreme ends of the range of levels of risk used in this experiment and therefore did not necessarily reflect the true ideal. Had the range of values been greater, most of these cases would undoubtedly not have exhibited equal ideals. The degree to which these data support the hypothesis can perhaps best be summarized in terms of the proportion of cases correctly predicted: In 82% of the cases where there was a difference in ideals, the direction of this difference could be predicted by knowing which pattern was preferred.

Table 1b shows the same sort of results for the comparison between Pattern 1/3 and Pattern 2/3. Again the hypothesis is strongly supported ($\chi^2 = 35.77$, $p < .01$), with correct

---

[4] This and subsequent indices are thoroughly described in the previous report (Coombs & Pruitt, 1960).

TABLE 1

CONTINGENCY TABLES SHOWING RELATION BETWEEN PATTERN PREFERENCE AND IDEAL LEVEL OF RISK

Ideal	Pattern preferred		
	$\frac{1}{3}$	$\frac{1}{2}$	Total
Ideal for $\frac{1}{3}$ greater than for $\frac{1}{2}$	15	4	19
Ideal for $\frac{1}{3}$ equal to that for $\frac{1}{2}$	13	17	30
Ideal for $\frac{1}{3}$ less than for $\frac{1}{2}$	8	40	48
Total	36	61	

a. $\frac{1}{3}$ vs. $\frac{1}{2}$

Ideal	Pattern preferred		
	$\frac{1}{3}$	$\frac{2}{3}$	Total
Ideal for $\frac{1}{3}$ greater than than $\frac{2}{3}$	29	4	33
Ideal for $\frac{1}{3}$ equal to that for $\frac{2}{3}$	12	6	18
Ideal for $\frac{1}{3}$ less than for $\frac{2}{3}$	8	33	41
Total	49	43	

b. $\frac{1}{3}$ vs. $\frac{2}{3}$

Ideal	Pattern preferred		
	$\frac{1}{2}$	$\frac{2}{3}$	Total
Ideal for $\frac{1}{2}$ greater than for $\frac{2}{3}$	34	5	39
Ideal for $\frac{1}{2}$ equal to that for $\frac{2}{3}$	12	18	30
Ideal for $\frac{1}{2}$ less than for $\frac{2}{3}$	12	15	27
Total	58	38	

c. $\frac{1}{2}$ vs. $\frac{2}{3}$

Note.—Entries are number of subjects.

prediction in 84% of the cases where there was a difference in ideals. Sixteen of the 18 equal cases were at an extreme end of the range of levels of risk. Table 1c, for the Patterns 1/2 and 2/3, also supports the hypothesis ($\chi^2 = 19.65$, $p < .01$), with correct prediction in 74% of the cases where there was a difference in ideals. All 30 equal cases were at an extreme.

Proposition 3. The more preferred a pattern of risk, the higher will be the maximum acceptable level of risk for that pattern.

Data concerning the maximum acceptable level of risk was available from the responses to Sets A, B, and C of the second experiment, where the subjects were asked to accept or reject any or all of seven levels of risk for three patterns, 1/3, 1/2, and 2/3. A level of risk was assumed to be acceptable to the subject if he accepted it at least five out of eight times it was presented to him. An index was derived for each pattern which ranged from 0 (no level of risk acceptable) to 7 (highest risk offered was the highest acceptable level). Preferences between the three patterns were ascertained in the usual way from the data of sets D, E, and F combined.

Proposition 3 was tested in the following form: If one pattern of risk is preferred to another, the maximum level of risk acceptable for the former will be greater than that acceptable for the latter. Data supporting this hypothesis are shown in Table 2.

In all three subtables there is a strong association between preference for a pattern and willingness to accept a higher maximum level of risk at that pattern, as hypothesized. (For Table 2a, $\chi^2 = 14.12$, $p < .01$; for Table 2b, $\chi^2 = 17.36$, $p < .01$; for Table 2c, $\chi^2 = 16.97$, $p < .01$.) Again the relatively large number of cases in which the maximum acceptable levels of risk for both patterns are equal do not necessarily weaken the hypothesis since most of them occur at the extremes of the scale (12 out of 17 in Table 2a; 7 out of 10 in Table 2b; and 13 out of 14 in Table 2c).

Proposition 4. For any given pattern of risk, divide the levels of risk into two groups: those smaller than the individual's ideal and those greater than his ideal. Within each group, levels of risk will be more preferred the closer they are to the ideal level of risk.

Consider the data from Sets I, II, and III of the first experiment. Proposition 4 implies that within each set, levels of risk above a subject's ideal will be preferred in order of ascending magnitude whereas levels of risk below his ideal will be preferred in order of descending magnitude, i.e., that the preference orderings will be folded **J** scales, in the terminology of the first report (Coombs & Pruitt, 1960). As shown in the first report, the results strongly supported this hypothesis: In Set I, out of 92 nonrandom orderings of preference among levels of risk, 81 were of the postulated type. In Set II, 89 out of 96 were of this type. In Set III, 83 out of 93 were of this type.

TABLE 2

CONTINGENCY TABLES SHOWING RELATION BETWEEN PATTERN PREFERENCE AND MAXIMUM ACCEPTABLE LEVEL OF RISK (MALR)

MALR	Pattern preferred		
	$\frac{1}{3}$	$\frac{1}{2}$	Total
MALR for $\frac{1}{3}$ greater than for $\frac{1}{2}$	6	2	8
MALR for $\frac{1}{3}$ equal to that for $\frac{1}{2}$	7	10	17
MALR for $\frac{1}{3}$ less than for $\frac{1}{2}$	0	14	14
Total	13	26	

a. $\frac{1}{3}$ vs. $\frac{1}{2}$

MALR	Pattern preferred		
	$\frac{1}{3}$	$\frac{2}{3}$	Total
MALR for $\frac{1}{3}$ greater than for $\frac{2}{3}$	12	3	15
MALR for $\frac{1}{3}$ equal to that for $\frac{2}{3}$	7	3	10
MALR for $\frac{1}{3}$ less than for $\frac{2}{3}$	1	13	14
Total	20	19	

b. $\frac{1}{3}$ vs. $\frac{2}{3}$

MALR	Pattern preferred		
	$\frac{1}{2}$	$\frac{2}{3}$	Total
MALR for $\frac{1}{2}$ greater than for $\frac{2}{3}$	13	2	15
MALR for $\frac{1}{2}$ equal to that for $\frac{2}{3}$	4	10	14
MALR for $\frac{1}{2}$ less than for $\frac{2}{3}$	1	9	10
Total	18	21	

c. $\frac{1}{2}$ vs. $\frac{2}{3}$

Note—Entries are number of subjects.

Proposition 5. For any given pattern of risk, all levels of risk up to the maximum acceptable level of risk will be acceptable.

Proposition 5 can easily be tested by an examination of the data from Sets A, B, and C of the second experiment. Within each set, the maximum acceptable level of risk is the highest level of risk which was accepted at least five out of eight times it was presented to the subject. The hypothesis is that within each set all levels of risk smaller than this maximum will also be accepted at least five out of the eight times they were presented. The hypothesis was confirmed in 110 of the 117 possible sets of data (3 sets times 39 subjects); which is, of course, exceedingly strong evidence for the proposition.

DISCUSSION

Comparison with Other Models

The four traditional models. The question naturally arises: Can the experimental results which support the *PLR* model be predicted or explained by one or more of the four traditional models? If so, one might argue that there is no need for a new model.

Two of these models yield clearcut predictions which are so grossly wrong that the models can easily be rejected in the context of these data. The *EV* model is obviously inadequate because it predicts indifference between the bets of these experiments (which all have an *EV* of zero) whereas in reality consistent preference was the rule. The *SEM* model is also obviously inadequate because, as mentioned earlier, it predicts that if a man likes a pattern at all he will insist on choosing the highest level of risk he can obtain for this pattern. Many subjects in the first experiment preferred moderate levels of risk, so that this prediction was not borne out.

The other two models are less easy to disprove because their predictions are not clearcut unless utility of money is carefully measured. Though such measurement was not attempted in these experiments, it seems reasonable to reject the models on another basis, namely that there is nothing in either the *EU* or *SEU* models, as usually stated, which would lead one to predict any of the regularities of behavior described in the five propositions and supported by the findings of these experiments. Perhaps, in the manner of the Procrustrean Bed, one of the traditional models could be stretched, new parameters added or more complex utility curves drawn to encompass these results. But the beauty of theory is its simplicity; so that it seems more reasonable to make striking new regularities of behavior such as these the basis of a new model.

Again it must be stressed that the traditional models should not be completely discarded; if possible, some common ground should be found between the new model and what is good about the old ones. In the first section it was shown that the model which best fit the segment of data to which it was applied was the *SEM* model (Edwards, 1955). A way in which this model can be incorporated into the *PLR* model without increasing the complexity of the latter will be shown in the section on research implications.

Probability and variance preferences. The *PLR* does not cast aside the concepts of probability and variance preference (Coombs & Pruitt, 1960; Edwards, 1953) but rather provides a larger framework within which they can be understood and related to one another. More specifically, the notion of "probability preference" is generalized in the concept "pattern preference," which embraces attitudes toward both probability and ratios of outcomes. The earlier finding (Edwards, 1954b) that probability preferences are unaffected by changes in money level is generalized into Proposition 1. The notion of "variance preference" is translated in the concept "ideal level of risk"; and the hypothesis that variance preferences fit a folded **J** scale of variance, proposed by Coombs and the author (1960), is expressed in Proposition 4, which is mainly dependent on the negatively accelerated character of the utility of risk curve. In addition, the *PLR* model goes a step beyond these earlier formulations by showing, in Proposition 2, how probability and variance preferences are related to one another and permitting predictions concerning the maximum acceptable level of risk in Propositions 3 and 5.

Research Implications

One of the important features of the *PLR* model is the implication of Proposition 1 that pattern preferences can be studied independently of level of risk. If true, this greatly simplifies research on gambling by permitting study of one dimension at a time; but this assertion must itself be further examined in the context of a wider range of patterns.

In studying pattern preferences, one might ask the question, What are the dimensions of pattern which determine preferences among them? Two dimensions which are suggested by the structure of simple wagers (e.g., "I'll give you 5 to 3 on the Cubs") are the subjective probabilities of outcomes and the ratios of objective outcomes to one another. For example, the utility of pattern in the *PLR* model could be expressed by a subjectively expected ratio (*SER*) model, which would be quite similar to the *SEM* model and would, therefore, hopefully retain its best features (prediction of choices between bets differing in pattern when level of risk is held constant). The formula for computing the utility of pattern would be:

$$u(p_X) = SER = \sum_i P_i^* t_i \quad [6]$$

where:

t_i equals $|\$_i| / \sum_i |\$_i|$,

the ratio of the (absolute) amount of money in the ith outcome to the sum of the (absolute) amounts of money in all of the outcomes of the bet.

One serious deficiency of the *PLR* model is its failure to predict to bets with no negative outcomes, where r_X equals zero. There is a very simple modification which would make it universally applicable (and satisfy the "sure thing" principle for bets with no negative outcomes), namely to add another objective parameter $m_X = \sum_i |\$_i|$ and rewrite the basic formula:

$$U(X) = m_X \cdot u(p_X) + g(r_X) \quad [7]$$

Since m_X is a monotonic, positive function of r_X for any pattern involving risk, Propositions 2 through 5 are unaltered by this procedure. Proposition 1 would have to be revised but not seriously. Since it has broader application, Formula 7 might have been used throughout this paper instead of Formula 5. However the latter is simpler and easier to present, and therefore was chosen.

REFERENCES

ALLAIS, M. Le comportement de l'homme rationnel devant le risque. *Econometrica*, 1953, **21**, 503–546.

BERNOULLI, D. Exposition of a new theory on the measurement of risk. *Econometrica*, 1954, **22**, 23–36.

COOMBS, C. H., & KOMORITA, S. Measuring utility of money through decisions. *Amer. J. Psychol.*, 1958, **71**, 383–389.

COOMBS, C. H., & PRUITT, D. G. Components of risk in decision making: Probability and variance preferences. *J. exp. Psychol.*, 1960, **60**, 265–277.

DAVIDSON, D., SUPPES, P., & SIEGEL, S. *Decision making, an experimental approach.* Stanford, Calif.: Stanford Univer. Press, 1957.

EDWARDS, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

EDWARDS, W. Probability preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, **67**, 56–67. (a)

EDWARDS, W. The reliability of probability preferences. *Amer. J. Psychol.*, 1954, **67**, 68–95. (b)

EDWARDS, W. Variance preferences in gambling. *Amer. J. Psychol.*, 1954, **67**, 441–452. (c)

EDWARDS, W. The prediction of decisions among bets. *J. exp. Psychol.*, 1955, **50**, 201–214.

EDWARDS, W. Behavioral decision theory. *Annu. Rev. Psychol.*, 1961, **12**, 473–498.

HURST, P. M., & SIEGEL, S. Prediction of decisions from a higher ordered metric scale of utility. *J. exp. Psychol.*, 1956, **52**, 138–144.

MOSTELLER, F., & NOGEE, P. An experimental measurement of utility. *J. polit. Econ.*, 1951, **59**, 371–404.

SNYDER, G. Deterrence and power. *J. conflict Resolut.*, 1960, **4**, 163–178.

SUPPES, P., & WALSH, K. A non-linear model for the experimental measurement of utility. *Behav. Sci.*, 1959, **3**, 204–211.

TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

(Received September 12, 1960)

Psychological Review
1962, Vol. 69, No. 3, 202–219

DISCRIMINATIVE CLASSICAL CONDITIONING IN DOGS PARALYZED BY CURARE CAN LATER CONTROL DISCRIMINATIVE AVOIDANCE RESPONSES IN THE NORMAL STATE[1]

RICHARD L. SOLOMON
University of Pennsylvania

AND

LUCILLE H. TURNER
Harvard University

The use of curare drugs to immobilize subjects in psychological experiments has special advantages for research on learning, transfer of training, and thinking. Subjects cannot give overt, skeletal responses, and therefore they cannot modify the exteroceptive or proprioceptive stimulus patterns that the experimenter wishes to control.

We shall show that a discriminative, classical conditioning procedure carried out on curarized dogs can later control instrumental avoidance responses in the normal state. We shall describe an experiment in which dogs were first trained to avoid shock by pressing a panel whenever a signal light (S^0) went out. The panel-press restored the light. Then, when the avoidance response was stable, the dogs were totally curarized. Under curare the dogs were presented on some trials with a tone (S^+) followed 10 seconds later by a shock of 5 seconds duration. A contrasting tone (S^-) was presented on other trials, never followed by shock. A few days later, after complete recovery from curare, the dogs were tested for avoidance responding to S^+, S^-, and S^0. All dogs then showed the differentiating effects of the classical, discriminative conditioning experience. The avoidance response to S^+ was usually of the same order of magnitude as that to S^0, while avoidance responding to S^- was weak, absent, or quick to extinguish. We do not believe that precise discriminative transfer of training—from the classical conditioning of reflexes to the instrumental learning of responses—has heretofore been demonstrated in this manner.[2]

It is of considerable theoretical importance, as well as of empirical interest, to know that skeletal avoidance responses acquired in the normal state can be controlled by classical conditioning procedures carried out under curare. Some current versions of S-R reinforcement theory emphasize the importance of differential peripheral, skeletal responses and their correlated proprioceptive cues in the development of discriminative instrumental acts.

---

[1] This research was facilitated by the Laboratory of Social Relations, Harvard University. Pilot studies over the past seven years were supported by grants from the Medical Sciences Division of the Rockefeller Foundation and a Ford Foundation Grant-in-Aid. The experiment in this paper was carried out with the support of a National Science Foundation Grant (NSF-G-14438). We are grateful to P. D. Watson, A. H. Black, N. J. Carlson, and M. R .Westcott for their assistance during many pilot studies.

[2] The general objective of our experiment certainly is not original, see Estes (1943, 1948). In these experiments rats were subjected to presentations of a tone followed by food, and then they were later trained to press a lever for food reinforcement. During a test period the rate of lever pressing was increased by presentation of the tone although the lever pressing response had not previously been associated with the tone. The major difference between Estes' experiments and ours lies in our use of the curare paralysis during discriminative conditioning so as to eliminate the possibility of response generalization from the conditioning situation to the instrumental learning situation.

Discriminative learning is hypothesized to take place by the action of differential reinforcement on two different responses to two different stimuli.

Because we shall show that discriminative instrumental acts can be established without those acts having been reinforced differentially (because they could not occur under curare), then a revision of theories emphasizing the importance of peripheral, skeletal responding and proprioceptive cues seems definitely to be required. Such a revision was once the goal of "latent learning" studies. But they failed to give clearcut evidence because the S-R reinforcement theorist who emphasized peripheral response components could always claim: (*a*) that overt responses were occurring during the latent learning period, (*b*) that some kind of reinforcement was apt to occur following those responses, and (*c*) that those responses were similar in some way to the final required performances. These were cogent objections. They can be effectively circumvented, however, by the use of the curare-conditioning preparation. Stimulus sequences presented to the totally curarized subject cannot be accompanied by differential skeletal responding. In addition, skeletal responding cannot occur in the presence of differential reinforcement. There are no overt acts for the experimenter to reward or punish. Thus the carry-over or transfer of "information" acquired under curare so as to influence skeletal responses in the normal state could be mediated *peripherally* only by differential ANS responding or "anxiety." Whether or not such a transfer phenomenon would be considered to be a function of "pure" central mediation or latent learning, or whether the mediation would be considered to be a function of anxiety, would, of course, depend on one's theoretical biases. A position emphasizing the mediation of learning and transfer by proprioceptive or kinesthetic cues suffers most in the light of our findings.

HISTORY OF PROBLEM

The theoretical issue to which our experiment is addressed is certainly an old one. In the days when psychology was dealing with mental content, there were those who maintained that the response processes of the skeletal musculature were necessary for perception and thought. In 1887, F. M. Mueller declared that we cannot think without words. This position was espoused by Watson and by Max when they identified implicit linguistic responses with thought processes. Early behaviorism was peripheralistic in nature. But so were the systematic approaches of many students of consciousness. William James emphasized peripheral response events in the production of emotional feelings. E. Jacobson maintained that awareness itself was a by-product of motor reactions. For example, he argued that muscular relaxation was accompanied by the disappearance of imagery and thinking.

While such peripheralistic interpretations of mental events were useful to the early behaviorists, the introspective methods used to arrive at the so-called "motor theory of consciousness" were shunned by them. At the hands of behaviorists, precise laboratory experimentation, aimed at the study of overt behavior as a function of stimulation conditions, replaced introspection. But the major idea of the "motor theory of consciousness—that the important mental processes such as perception, feeling, thinking, and imagining were *caused by* peripheral motor reactions—was retained by the peripheralistic behaviorists in new form. Complex behavioral phenomena, difficult to place in the simple S-R formula, were interpreted in a peripheralistic way. Thus, animals did not think, but instead they showed transfer of training *mediated* by chains or organizations of "response-produced cues." Animals did not "expect," but instead they maintained their future-orientations by virtue of response-produced cues which served to bridge time. Implicit anticipatory goal responses became a major conceptual tool of the Hullians, while differential proprioceptive and kinesthetic cues carried the Skinnerian subjects over their toughest problem solving challenges. Of course, many of the old theoretical controversies revolved around the validity of these peripheralistic positions.

Harlow and Stagner (1933), following a

suggestion from Donald G. Marquis, were the first to see the importance of a paralyzing drug in hitting at the crux of the peripheralist-centralist controversies, whether the controversies concerned the motor theory of consciousness or the latent learning concept and its challenge to peripheralistic learning theory. These experimenters argued that, if the motor theory of consciousness were correct, conditioning in paralyzed animals should not transfer to the normal state, since the conditioning experience would be ineffective. They further pointed out that if discrimination learning were a function of the differential reinforcement of overt skeletal acts, then conditioning in paralyzed animals should not transfer to the normal state, since there would be no differential responses in the paralyzed state.

Harlow and Stagner used the drug curare in their experiments. Curare produced a complete flaccid paralysis in animal subjects. Its action was at that time thought to consist mainly of chemical blocking at the junction of the motor axon and muscle cell. So curare could remove skeletal motor responses and their proprioceptive feedback and create a preparation in which both the motor theory of consciousness and the theory of latent learning could be tested. Therefore, the experiments by Harlow and Stagner (1933) were considered to be crucial. Harlow and Stagner studied both instrumental avoidance learning and pupillary conditioning in curarized cats and dogs. In one experiment, they first pretested subjects for aversion to the following stimuli in the experimental room: a chair, bell, light, and buzzer. Then the subjects were put on a metallic grid, and the latency of jumping off it, in response to the various pretested stimuli, was measured. Finally, all subjects were curarized to the stage where there was "absence of muscle twitches, gag reflex, and the corneal eye-lid reflex" (Harlow & Stagner, 1933, p. 287). Then followed 30 pairings of conditioned stimulus and electric shock to the grid on which the subjects had been placed. Under such treatment a normal, undrugged control group learned to jump off the grid in response to the conditioned stimulus (CS) in three trials or less. When the curarized subjects had recovered from the effects of the drug, they were reintroduced to the grid, presented with the CS, but showed not the slightest evidence of learning.

In a second experiment, Harlow and Stagner presented in sequence to curarized subjects a light, a bell, and a shock. The pupil initially contracted to the light, dilated to the shock, but showed no change to the bell. But after two to six presentations of the three-stimulus sequence to curarized subjects, the pupillary dilation, originally elicited by shock, became elicitable by the bell alone. Said the experimenters (Harlow & Stagner, 1933):

> It thus appears certain that cortical activity (to the extent required for the acquisition of the conditioned pupillary response) is not depressed by curare, at least in the amounts used in this experiment (p. 289).

Then two extinction experiments were carried out. Three cats were trained to leap from the electrified grid in response to a CS. Later, when curarized and placed on the grid, they were presented with the CS alone. When the subjects recovered from curarization, and were placed on the grid again, they leaped from the grid with short response latencies. In another group, pupillary dilation was conditioned while the subjects were in the normal state. Then, while curarized, the CS was presented from 100 to 150 times without the shock, and the conditioned dilation disappeared. Yet skeletal avoidance responses were later shown to have remained intact in these same subjects. Harlow and Stagner felt that they had established the fact that adaptive behavior is not learned under curare. However, they asked whether it might be possible that CS-US pairings under curare might have central effects which aren't organized in such a way so as to produce a specific response. A savings test might reveal this. Therefore, two control puppies were given 30 CS-US sequences under curare, then trained again later in the normal state. They were compared to two puppies which were trained only in the normal state. There were no differences between the two pairs of subjects in either acquisition or extinction rates. In another experiment, conditioned forelimb flexion was used (more limited and segmental response than jumping off the grid), but it revealed no learning under curare. The conclusion reached by Harlow and Stagner (1933) was this: ". . . it appears that presentation of stimuli alone will not cause learning if no reaction is made" (p. 293). The conclusion hit hard at latent learning theory.

In direct contradiction to the opinion of Harlow and Stagner, Light and Gantt (1936), using a different method, concluded that ". . . the peripheral nerve and executor organ are not necessary for conditioned reflex formation" (p. 35). In four dogs the right hind leg was paralyzed by crushing the anterior nerve roots. Before regeneration of the injured nerves, simple conditioning was carried out on the paralyzed side. A CS was paired with shock to the immobile leg; when a diffuse conditioned response (CR), consisting of howling and struggling, was established, training was stopped. Later, after regeneration of the damaged motor nerves, the CS was presented, (Light & Gantt, 1936) "and it was followed by withdrawal of the formerly paralyzed leg—the appropriate and specific conditioned movement, but one which was never possible during the period of training" (p. 35). At the time it was certainly possible that the Harlow and Stagner experiment was not inconsistent with this result. After all, Light and Gantt had used a classical conditioning procedure; and this procedure had produced pupillary conditioning in the immobilized state for Harlow and Stagner. (The psycho-pharmacological properties of the curare used by Harlow and Stagner had not yet come under scrutiny.) Then, too, the Crisler (1930) findings had indicated that classical conditioning could occur in the case of a blocked salivary reaction. Perhaps the distinction that would resolve the discrepancy was that between classical conditioning and instrumental training. Perhaps the conclusion of Harlow and Stanger applied only to the latter procedure?

Settlage (1936), using a Pavlovian delayed conditioning procedure (CS-US interval of 2 seconds), attempted to condition 12 cats under sodium amytal. The drug level was adjusted so that when the CS was presented, the subjects gave no anticipatory limb withdrawals. Then conditioning pairings of CS and shock were given, ranging from 10 to 50 in number from stimulus to stimulus. Later, in the normal state the CS was presented alone. A high percentage of limb withdrawals was noted. Settlage felt that a new S-R connection was established under sodium amytal. However, this finding was not critical for a peripheral-motor learning position because, while under amytal the unconditioned stimulus (US) was often capable of eliciting a limb withdrawal, even though the CS was not. It was, however, an interesting and instructive finding. Here was a case where the organism was in a state such that an unconditioned response (UR) could be elicited yet no CR could emerge; however, the CR was later shown to have developed normally under sodium amytal immobilization, since it occurred when the CS was presented to the animals after they had recovered from drug effects.

The psycho-pharmocological complications of the use of curare were now to come under scrutiny in a puzzling series of experiments, many by Girden and his collaborators. The first of these (Girden & Culler, 1937) demonstrated certain disturbing side effects of the drug on the function of the brain. The semitendinosus muscle of the dog was dissected out, leaving its circulation and innervation intact. Then the subject was presented with a 2-second bell followed by a 0.5-second shock to the paw of the prepared leg while the subject was totally curarized. However, the US did elicit a twitch of the semitendinosus muscle, even though it did not produce a limb movement. A good semitendinosus muscle twitch CR developed in response to presentation of the CS under curare. Yet this CR vanished when the dogs were allowed to recover from the effects of curare. The CR reappeared when the dogs were curarized again. When a group of dogs was conditioned in the normal state, then tested under curare, there was no detectable CR of the semitendinosus muscle. Yet the CR reappeared when the effects of curare wore off. There thus was a lack of transferability of conditioning from one state to the other. Girden and Culler (1937) attributed the findings to a separation of cortical and subcortical conditioning processes; they said:

> It is thus conceivable that under curare the normal cortical dominance is inhibited, and that conditioning therefore occurs at subcortical levels (thalamus). When the animal revives, the cortex again functions normally and the (conditioned) thalamic activities are inhibited. Likewise the CR established in the normal animal (with participation of the cortex) is inhibited under curare (due to general inhibition of the cortex) (p. 272).

Whether or not this interpretation would endure, it was, however, certain that the curare used in these experiments produced complicated central nervous system effects; it was not merely a peripheral-motor blocking agent.

Harlow and Settlage (1939) designed an experiment to detect the effect of curarization of the upper body upon the ability of cats to display previously learned responses of the lower body. The subjects' aortas were prepared so that they could be tied off in the abdominal region. Then the subjects were conditioned to give vigorous hind leg withdrawals, using a shock as US, and a bell as CS. After conditioning, the aortas were clamped off and curare was injected into the upper body. The upper body parts became paralyzed, while the lower body parts could move until they became cyanotic due to the occluded blood supply. Vigorous CRs persisted for short periods, perhaps no longer than 3 minutes after curarization. It was not possible to tell whether the disappearance of CRs was due to collateral circulation of curare to the legs or to lack of adequate blood supply. The experimenters felt that they had demonstrated that curare does not immediately affect the CNS, at least not for 3 minutes following injection.

Culler, Coakley, Shurrager, and Ades (1939) bared the spinal roots and the cerebral cortices of nine dogs. They located the spinal root locus and cerebral cortex locus for electrical elicitation of twitches of the semitendinosus muscle. Then they curarized the dogs and measured the stimulation intensities necessary to produce a constant magnitude of excursion of the semitendinosus muscle. They found that when they compared these stimulation intensities for curarized compared to normal dogs, a striking result emerged. Curare elevated the excitation time as well as the rheobase for the cortex, while there was almost no effect, and perhaps the opposite, at the ventral root. The experimenters felt that somewhere between the cortex and the ventral spinal root lies a plane of neural cleavage above which curare is a depressant. While this might explain the findings of Girden and Culler (1937), it still did not agree accurately with the observations of Harlow and Stagner (1933) on the successful development of pupillary conditioning under curare, unless one would assume further that pupillary conditioning occurs below the postulated plane of neural cleavage.

Harlow accepted the conclusion that curare was a CNS depressant. However, he felt that if dosage could be adequately adjusted, one might produce sufficient immobilization without severe CNS side effects. Therefore, in an extensive series of experiments, he employed what he called "incomplete curarization." Using 11 cats that were incompletely paralyzed, he established conditioned escape responses (with a 4-second CS and 1-second shock) with some difficulty. In contrast, conditioned pupillary reactions developed quickly. Later in the normal state, the conditioned escape responses were vigorous, as were the pupillary reactions. Harlow felt that curare is a CNS depressant; but in light doses it permits transferable CRs to be formed. He (Harlow, 1940) further stated that a failure of learning to proceed normally under curare was due to the cerebral depressant effects. "Thus the data from the curare experiments cannot be used to support the motor theory of learning, and clearly refute such a theory if it is to be applied to all learning phenomena" (p. 281). This conclusion was not very convincing when applied to learned skeletal responses. After all, Harlow's subjects had been able to perform escape responses during the curarization period.

Then followed as persevering and careful a series of experiments as one can find in modern experimental psychology. Over a span of years from about 1938 to 1947, Edward Girden elucidated the behavioral properties of the curare drugs then available (curare and erythroidine); our knowledge of "dissociation" under curare comes from his work. Girden (1940) performed a bilateral surgical extirpation of the cortical auditory areas of dogs. After recovery from the aftereffects of surgery, the semitendinosus muscle was exposed. Then the dogs were subjected to a conditioning procedure identical to that used in Girden's 1937 experiment (Girden & Culler, 1937), using an auditory CS and shock US. Dogs conditioned while curarized gave good CRs later in the normal state; and dogs conditioned in the normal state gave CRs later when curarized. This outcome was in direct contrast to that of the earlier experiment. Here is what Girden (1940) thought the results meant:

> It is clear from the present results that conditioned responses of the semitendinosus muscle to an auditory stimulus in the normal animal are mediated by cortical pathways, whereas the CRs developed under curare involve subcortical systems. When a section of the cortical pathway is literally extirpated, the block between the normal and curare states is disrupted; the CRs established in either of the two conditions will now be manifested in the other state (p. 404).

Girden voiced the conviction that *normal animals conditioned under curare are probably unconscious.* Furthermore, he (Girden, 1940) entertained the idea that even in his dogs with removed auditory cortex the same condition might have existed.

> In the partial decorticate, the amnesic condition is disrupted so that the curare CR now appears in the subsequent normal period. But the animal still appears to be unaware of what had occurred under the drug. That is, the animal does not manifest the diffuse struggle behavior ordinarily present during the first stages of conditioning. . . . It seems as if the animal is still unaware of what had happened under the drug, the semitendinosus CR occurring much like the motor automatisms reported in hysterics (pp. 405–406).

If this conclusion were true, then the usefulness of curare, for the investigation of learning without skeletal responding, would certainly be nil.

Girden (1942a) used intact dogs rather than the excised semitendinosus preparation. In addition to the old curare drug, he used erythroidine, a variant of curare. Some subjects were conditioned to give defensive limb withdrawals to a CS while in the normal state. They were later tested in the drugged state and showed no retention. Some subjects were conditioned by CS-US pairings while drugged and were later tested while normal. These showed no retention either. Even when the dosages of curare or erythroidine were light, permitting gross bodily responses to the US during conditioning in the drugged state, there was no retention in the normal state. This dissociation phenomenon was likened to amnesia.

Girden (1942b) then extended his studies to the conditioning of visceral reactions. A glass canula was inserted in the common carotid artery of the dog for purposes of blood pressure recording. Using a light flash as CS and shock to forepaw as US, and paralyzing the dogs with erythroidine, Girden showed that a blood pressure increase could be readily conditioned, extinguished, and reconditioned. However, the retention of a blood pressure CR in the normal state was not measured. On this omission, Girden (1942b) said: "It is safe to infer, however, that the blood pressure CR, developed in the drug state, was suppressed after recovery as were the other components of the drug state CR which have been studied" (p. 230). This conclusion was based on the fact that Girden's dogs, after recovery from drug effects, did not respond to the CS with startle, respiration changes, or pulse rate changes.

In view of the fact that Harlow and Stagner (1933) had believed that the conditioned pupillary reflex, established under curare, was retained later in the undrugged state, the question of dissociation for visceral or autonomic CRs was certainly not settled yet. Therefore, Girden (1942c) returned to the study of the conditioned pupillary reaction. In dogs immobilized by either curare or erythroidine, a conditioned pupillary dilation was established, using a bell CS and shock to forepaw as US. The conditioned reaction was shown to be due to sympathetic innervation, and not due to inhibition of parasympathetic innervation, since this CR failed to develop if sympathetic fibers to the pupil were cut. The pupillary CR was readily extinguishable in the drug state. But the CRs established in the drug state did not transfer to the later normal state. So Girden (1942c) stated: "We cannot concur, therefore, with the conclusion that 'curare produces different influences upon the various components of a total conditioned response pattern'" (p. 331). He was in complete disagreement with Harlow and Stagner (1933) on this point. In addition, Girden, in order to maintain that the dissociation phenomenon was due to a block between cortical and subcortical functions, would have had to assume that pupillary conditioning in the normal state was cortically mediated. He was, however, silent on this point at the time. Later, Girden (1943a) experimented with different levels of erythroidine dosage, producing mild to very deep paralysis in dogs. He found that autonomic CRs were established under erythroidine but were not detectable later in the normal state, no matter how deep the flaccid paralysis of striate muscles had been. He further found that only those striate muscle reactions actually made under erythroidine, can become conditioned. When they do, they are part of a generalized "excited" emotional reaction pattern. Girden conjectured that a human subject, rendered partially immobile by erythroidine, would be unconscious; but later on in the normal state, any autonomic CRs established in the drug state would be rearoused. The subject would perceive his own autonomic upset without knowing how it had developed. Such autonomic CRs would only transfer to the normal state if they were artifacts of a general skeletal emotional response pattern which had occurred under light paralysis. Thus Girden tried to tie together autonomic and skeletal CRs into an emotional reaction pattern. The autonomic reactions are side effects of the skeletal emotional responses. With deep paralysis, therefore, autonomic CRs would not later transfer to the normal state.

In the same year, Girden (1943b) showed that the EEG in rhesus monkeys and dogs remained normal under flaccid paralysis induced by erythroidine. This was true even after repeated dosages of erythroidine had immobilized subjects for as long as 4 hours. This finding put in doubt the role of cortical depression in the dissociation phenomenon. But it did show that artificial respiration, required in deeply curarized subjects, was adequate for normal brain function in this and previous experiments in Girden's laboratory. This was no trivial finding, in view of the peculiar amnesias found with curare immobilization. Finally, Girden (1947) extended his studies to rhesus monkeys. Using erythroidine, four subjects were avoidance trained under light dosages. Using both visual and acoustic CSs, CRs rapidly developed under erythroidine. Girden (1947) reported that such CRs often developed more rapidly in the drug state than in the normal state.

> As previously observed in the dog, there is a functional dissociation in the curarized monkey with the result that drug state learning is completely repressed upon complete recovery from the drug. The CR appears spontaneously when the monkey is again curarized (p. 587).

The development of autonomic CRs was normal under erythroidine, but there was no transfer to the normal state. We wish to emphasize the striking finding that discrete, localized skeletal CRs, as well as autonomic CRs, were found to develop faster in the drug state than in the normal state. Girden had previously noted this in dogs, and had commented upon it. In one study, using the semitendinosus response, 17 subjects had reached an arbitrary conditioning criterion under curare in 23 to 125 trials. In the normal state, another group of subjects had required 100 to 250 trials to reach criterion. The observations on monkeys confirmed this finding.

Morgan, (1951) in tracing the history of the use of curare and erythroidine in behavioral experiments, accepted Girden's general conclusions (Morgan, 1951): "Thus the story is now simple and clear. Conditioning reactions, whether skeletal or autonomic, established under curare or erythroidine, do not carry over to the normal state and vice versa" (p. 771). Clearly these two drugs could not be legitimately used for the original purpose of the Harlow and Stagner experiment. If one wished to study the importance of the peripheral skeletal response in learning, other immobilizing techniques would be required.

It turned out that Girden stopped his experimentation on curare drugs too soon. A new curare derivative, d-tubocurarine, had been administered to a human subject for experimental purposes in 1945. The object was to test its efficacy as an anesthetic. For several years prior to this time, both curare and erythroidine had been used as an anesthetic supplement for surgery. Sometimes they had been used alone for surgery. They were effective muscular relaxants, and as such were an aid to surgical procedures. However, when d-tubocurarine was used alone as an anesthetic, sometimes the patient would complain after the operation that he had experienced severe pain. This had not happened with curare or erythroidine. Therefore, Smith, Brown, Toman, and Goodman (1947) used the new paralyzing drug, d-tubocurarine, on Smith to see whether it had anesthetic or cerebral depressant properties. Smith was given a dose large enough to induce total flaccid paralysis. He was artificially respirated. The other three experimenters presented their colleague with a wide variety of external stimuli, some painful, while he was immobile. After the subject recovered he was able to relate in accurate detail most of what had transpired. He said his "senses" were extremely "clear" under d-tubocurarine. His blood pressure remained normal during the procedure, and except for the discomfort of needing to cough and being unable to do so, suffered no ill effects. This outcome, the authors stated, agrees with the finding of Girden (1948) that curare drugs do not depress the EEG as long as artificial respiration is adequate. It is not clear why Smith et al. (1947) ignore Girden's work on dissociation while accepting that on the EEG. They state that Girden, employing CR techniques with animals, believed his results to indicate that curare causes unconsciousness or amnesia; but they add that his evidence is not very conclusive. On the contrary, we would feel that his evidence was good, and it agreed with the clinical findings on the old curare as an anesthetic agent. Clearly, d-tubocurarine was quite different from curare and erythroidine in its psychological action.

The Smith et al. (1947) findings have been partially confirmed by perceptual tests with human subjects. McIntyre, Bennett, and Hamilton (1951) found that doses of d-tubocurarine left CNS processes unaffected. Unna and Pelikan (1951) found, with human subjects (six healthy volunteers) that small, subparalytic doses of d-tubocurarine did not alter a wide variety of perceptual functions. They (Unna & Pelikan, 1951) stated: "No evidence was obtained of any action other than on the neuromuscular junction. . . . In particular, no effects on autonomic organs and also none on cerebral functions could be demonstrated . . ." (p. 480). Lauer (1951) demonstrated the transfer of learned forelimb flexion in two dogs from the immobile state to the normal state, using d-tubocurarine to produce complete paralysis. Beck and Doty

(1957) have confirmed the same type of finding in cats, using a cataleptic compound, bulbocapnine. In an extensive monograph now in press, Black, Carlson, and Solomon (1962) have demonstrated several related phenomena of autonomic conditioning in dogs paralyzed by curare. They have confirmed the fact that the more recent curare drugs (flaxedil, d-tubocurarine) produce a complete flaccid skeletal paralysis *without the dissociative or amnesic effects produced by the curare drugs* used previously in the pioneering work by Harlow and Stagner (1933), Girden (1940, 1942a, 1942b), and others. The classical conditioning of ANS reactions in curarized dogs has been shown to transfer to conditioned reflexes in the normal, undrugged state. In addition, the development of ANS discriminative, conditioned reactions has been demonstrated in curarized dogs. This means that ANS conditioning is not merely an artifact of skeletal responding or its proprioceptive feedback; it means also that the original experimental idea of Harlow and Stagner can now be carried out as a satisfactory test of the "central process" versus "peripheral process" interpretations of learning and transfer.

METHOD

Subjects

Subjects were six healthy mongrel dogs obtained from the Harvard Medical School animal farm. They were of medium size (10-12 kilograms). Their previous histories were unknown because they were strays.

Apparatus

The apparatus has been described in some detail by Black (1958). A Pavlov-type harness held the subject in a comfortable, suspended posture. Two panels were placed on either side of the subject's head. The stimulus light (S^0) was placed 1 foot above and 4 feet in front of that subject's head. The light was provided by a 40-watt bulb. The tones (S^+ and S^-) were provided by two Bud Code-practice oscillators, set at approximately 160 cycles per second and 1200 cycles per second. They were placed 4 feet in front of, and 2 feet above the dog's head. The amplitude was set at a level that seemed nonaversive to the experimenters.

Appropriate interval timers, relays, and programming equipment were used to control the CS-US interval, US duration, and intertrial intervals. The accuracy of the CS-US interval timer was 10 seconds ± 0.05 second. The accuracy of the shock duration timer was 5 seconds ± 0.02 second.

The EKG was recorded on a Grass four-channel polygraph. Alligator clip electrodes, attached to shaved skin areas, were used to pick up EKG impulses. The shock electrodes were those described by Black (1958); they were brass plates embedded in Bakelite, and they were laced tightly to the hind toe pads.

Procedure

Stage one—avoidance training. The subjects were placed in the harness, EKG and shock electrodes were attached, and the panels were adjusted so that each dog could readily press either panel. The experimenters then left the experimental room. Lights were turned off (with the exception of the signal light and one dim 5-watt lamp which hung above and behind the subject). The signal light shone directly on the subject's face. In the door to the experimental room was a one-way mirror through which the experimenters observed the subjects.

When the initial struggling period seemed over and the subject was calm, training was begun. At least three test trials were given at the outset, in order to be sure that there was no initial panel-pressing response to the S^0 (light out). Then the S^0 was paired with shock, with a CS-US interval of 10 seconds. Shock stayed on until the subject pressed either panel, at which moment S^0 and shock were simultaneously terminated. The intertrial intervals were either 1, 1½, or 2 minutes, presented in random order. Stimulus presentation, trial presentation, recording of responses, as well as control of the S^0 and shock onset, were automated so that the subject could be left unattended. If the subject pressed a panel in less than 10 seconds following CS onset, then S^0 was instantaneously removed and no shock was presented. The subjects were trained until they gave reliable avoidance responding. Pressing either panel prevented shock. A criterion of at least 20 consecutive avoidance responses was required, but each subject was trained to a different criterion. The differing avoidance criteria were used in order to vary the number of trials since the last shock was received prior to curare-conditioning. Then the subjects were tested for sensitization by presentations of the two stimuli which were to be used in the next stage of the experiment. A 160 cycles per second tone and a 1200 cycles per second tone were presented three times each, in a random order.

Stage two—conditioning under curare. After the subject met the avoidance criterion, he was curarized. D-tubocurarine was administered i.e., with a dose of at least 0.7 cubic centimeter per kilogram, or 2.0 milligrams per kilogram, needed to ensure total flaccid paralysis. A gradual injection procedure was used, starting with an initial dose of 6.0 milligrams, and thereafter injecting 0.3 milligram per minute. When respiration stopped, usually 30 seconds after an initial 6.0 milligrams dose, an endotracheal tube was inserted, and artificial respiration was begun.

Conditioning was started when the subject showed no respiratory movements, no corneal lid reflex, no ear or tongue twitches, and no paw movements. Conditioning trials were carried out in the Pavlovian manner. A delayed conditioning procedure was used with a 10-second CS-US interval and a shock of 5 seconds duration and of 4 milliamperes intensity. Thus the CS lasted 15 seconds, and the US occurred during the last 5 seconds of this interval. The CS was terminated at the moment of shock termination. Two stimuli were used: S^+, a tone paired with shock, and S^-, a tone never paired with shock. The 160 cycles per second tone was S^+ for dogs number 21, 22, 23, and 25. The 1200 cycles per second tone was S^+ for dogs number 32 and 33.

All six subjects were given 99 curare-conditioning trials. The following sequence, $+ + + - + + - + - + - - + - - -$, was repeated five times and then was followed by a sequence of 19 alternated + and − trials, ending on the 99th trial with a + trial.

Stage three—transfer tests. After the curare-conditioning session, all subjects were given at least 48 hours to recover from the side effects of curarization. Then they were tested in the normal state in the same apparatus as that used previously for both avoidance training and curare-conditioning. All three CSs were presented, and the behavior of the six subjects in response to these stimuli was recorded. *No shocks were given, so the test sessions were extinction sessions.*

The order of test presentations of the three types of CSs presented serious problems of potential biasing. While there were no shocks presented on these test trials, there was the possibility that order effects could either enhance or diminish the panel-pressing responses to the three CSs. For example, if the subject were at first presented with a series of S^- presentations and did not respond, there was the possibility that responses to the S^+ or S^0 might become less probable due to some type of generalization of inhibition. Or, perhaps, if a long series of S^0 trials were presented, this might facilitate responding to S^- and S^+ due to some type of generalization of excitation or a sensitization (or pseudoconditioning) phenomenon. We were not certain how to solve this problem without a very large number of subjects. Therefore, we explored different orders of CS presentation with different subjects. The purpose was to eliminate biasing for each subject, in order to determine whether the three CSs had differential elicitation strengths.

Latencies of the panel-pressing responses were measured from CS onset to the time the panel was depressed 1 inch. The CS was arbitrarily terminated if the subject did not press the panel in 15 seconds (the time interval initially required for the 10-second CS-US interval plus the 5-second shock duration period). If no response was made in 15 seconds, it was recorded as an infinite latency. It was known from previous studies that the termination of the CS after a short time period facilitates extinction. Therefore, we were using conservative conditions for the demonstration of response differentiation.

RESULTS

In Table 1 are listed the total number of shocks received by each subject during avoidance training and the number of consecutive avoidance re-

TABLE 1

SHOCKS RECEIVED, AND RESPONSES TO REACH CRITERION, FOR ALL SUBJECTS

Subject's code number	Name of subject	Number of shocks received	Consecutive avoidance responses to criterion
21	Harry	46	25
22	Bugeye	43	58
23	Buff	88	20
25	Nipper	52	20
32	Punc	119	77
33	Mahog	267	88

sponses made in reaching the arbitrary avoidance learning criterion.

The latencies of panel-pressing are the more important observations of this experiment; these are plotted in Figures 1–6 for the last 20 training trials prior to curare-conditioning and for the test trials after recovery from curare effects. The data on sensitization tests are not plotted since all these pretest responses to S^+ and S^- were of infinite latency. In Figures 1–6 responses to S^0 are indexed by filled circles; responses to S^+ are indexed by closed triangles, and responses to S^- are indexed by open triangles. On the ordinate is the reciprocal of latency (× 100) on a logarithmic scale. The arbitrary infinite latency (no response in 15 seconds) is at the bottom of the ordinate scale.

It is convenient for us to describe our results by individual subjects. *All subjects lead us to the same general conclusions but show slightly differing phenomena.*

Subject #21, shown in Figure 1, on the last 20 trials before curare-conditioning, showed normally stable avoidance responding to the original training stimulus, S^0. When tested on S^+ and S^- before the curare-conditioning trials, he gave no responses. Like most subjects, he pricked up his ears, looked around, but made no incipient panel-presses. After curare-conditioning, we can see that there still was no response made to the S^-. On the other hand, a few responses appeared to S^+, and these were of short latency. The latency of responses to S^0 were of the same order of magnitude as were those

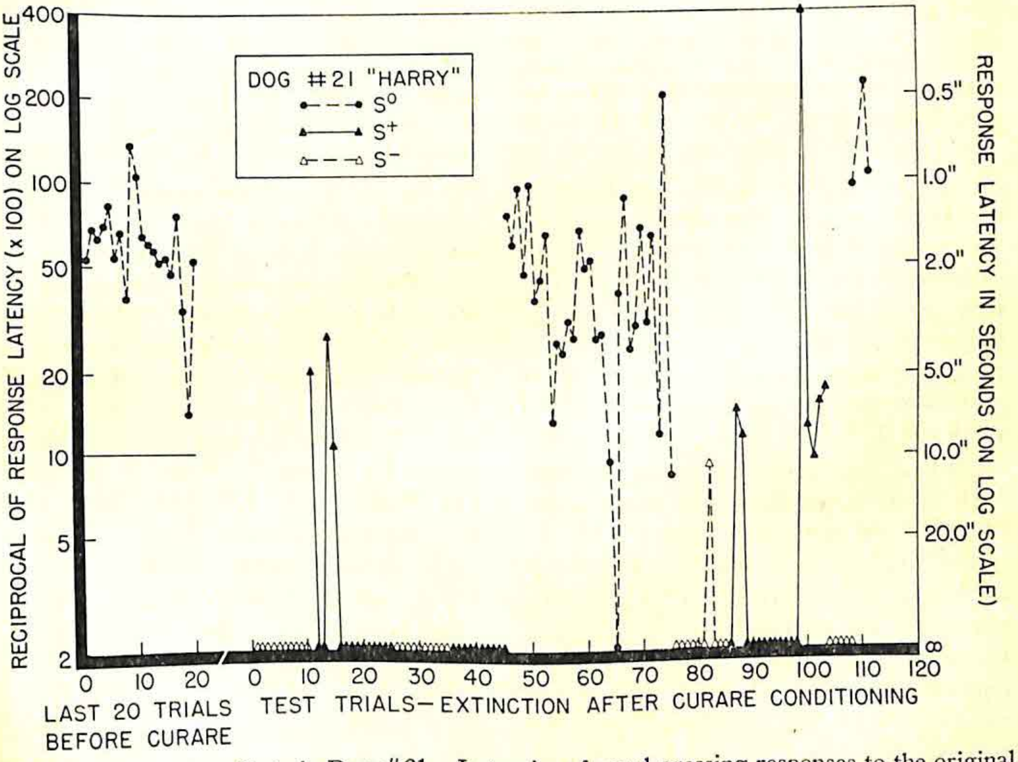


FIG. 1. Transfer effects in Dog #21. Latencies of panel-pressing responses to the original training stimulus, S^0, and to the two stimuli, S^+ and S^-, used in the curare-conditioning procedure. (Note that the latencies are expressed in reciprocal form. Note that this dog shows very little differentiation of S^+ and S^-.)

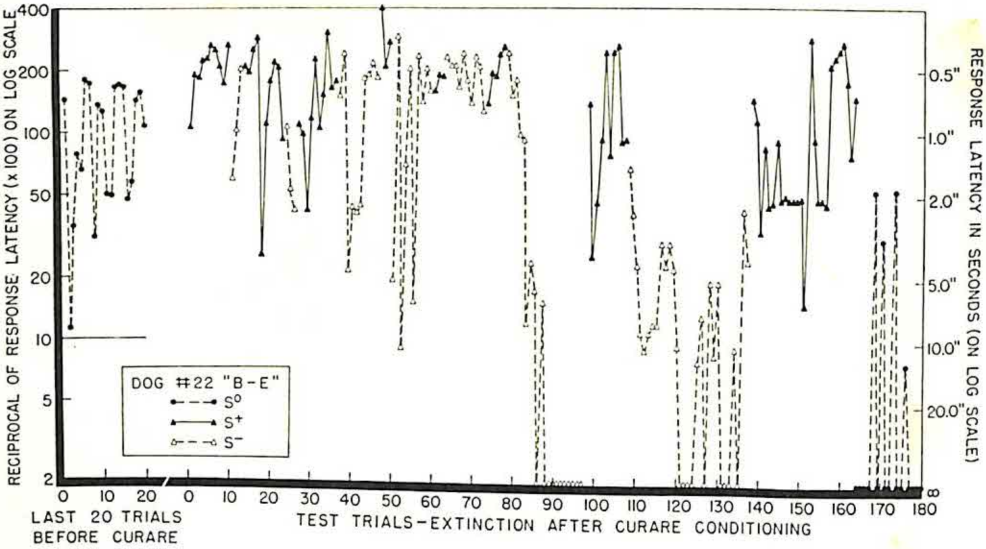


Fig. 2. Transfer effects in Dog #22. Latencies of panel-pressing responses to the original training stimulus, S^0, and to the two stimuli, S^+ and S^- used in the curare-conditioning procedure. (Note that the latencies are expressed in reciprocal form. Note that this dog shows very little differentiation of S^+ and S^- on early test trials, but there is clear evidence of more rapid extinction of panel-pressing to S^- than to S^+.)

prior to curare-conditioning. One response was made to S^- in 35 tests. Ten responses were made to S^+ in 42 tests. We felt that the presentation of 10 S^- tests at the outset inhibited the responding to S^+ on later tests. We were especially struck by the fact that 30 subsequent test trials on S^0 led to a facilitation of panel-pressing to S^+. Following these 30 S^0 trials, there were 7 responses in 18 S^+ trials and 1 response in 15 S^- trials. The results on this animal were promising enough to lead us to replicate the experiment on other subjects.

Subject #22, shown in Figure 2, achieved stable avoidance responding prior to curare-conditioning. When tested with S^- and S^+ prior to curare-conditioning, he showed no panel-presses, indicating no sensitization effect at that stage of the experiment. Following curare-conditioning, however, this subject pressed the panel in response to all three CSs. But after approximately 40 test trials, the latencies of avoidance responding to S^- were clearly more variable than those to S^+. After approximately 80 test trials, the response latencies to S^- were significantly longer than those to S^+, and there were several failures to respond to S^-. If we look at Test Trials 110–160, there can be no doubt that Subject #22 was responding strongly to S^+ but very poorly or not at all to S^-. There were nine failures to respond to S^- on Test Trials 110–137. On Test Trials 138–163, there were no failures to respond to S^+. Interestingly enough, when finally tested on S^0, on which Subject #22 had originally been trained, there were 13 failures to respond in 17 tests.

Subject #23, shown in Figure 3, gave stable avoidance responding to S^0 prior to curare-conditioning. When tested for sensitization on S^+ and S^- prior to curare-conditioning, he gave no responses. Following curare-con-

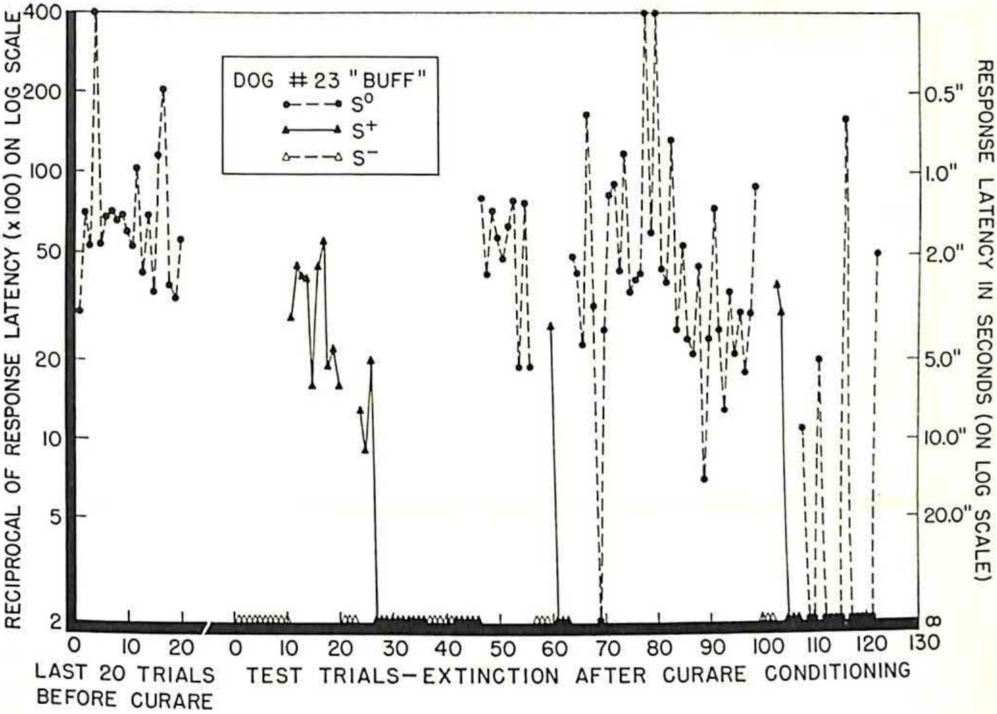


FIG. 3. Transfer effects in Dog #23. Latencies of panel-pressing responses to the original training stimulus, S°, and to the two stimuli, S⁺ and S⁻ used in the curare-conditioning procedure. (Note that the latencies are expressed in reciprocal form. Note that this dog shows excellent differentiation of S⁺ and S⁻ at the beginning of transfer tests, and that responses to S⁺ extinguish rapidly.)

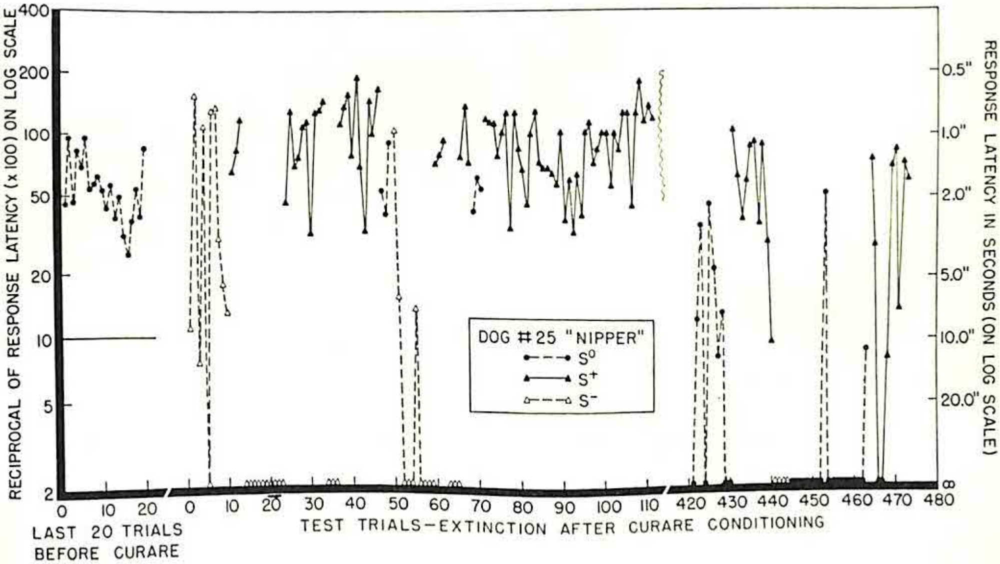


FIG. 4. Transfer effects in Dog #25. Latencies of panel-pressing responses to the original training stimulus, S°, and to the two stimuli, S⁺ and S⁻ used in the curare-conditioning procedure. (Note that the latencies are expressed in reciprocal form. Note that the responses to S⁻ extinguish quickly while those to S⁺ remain stable.)

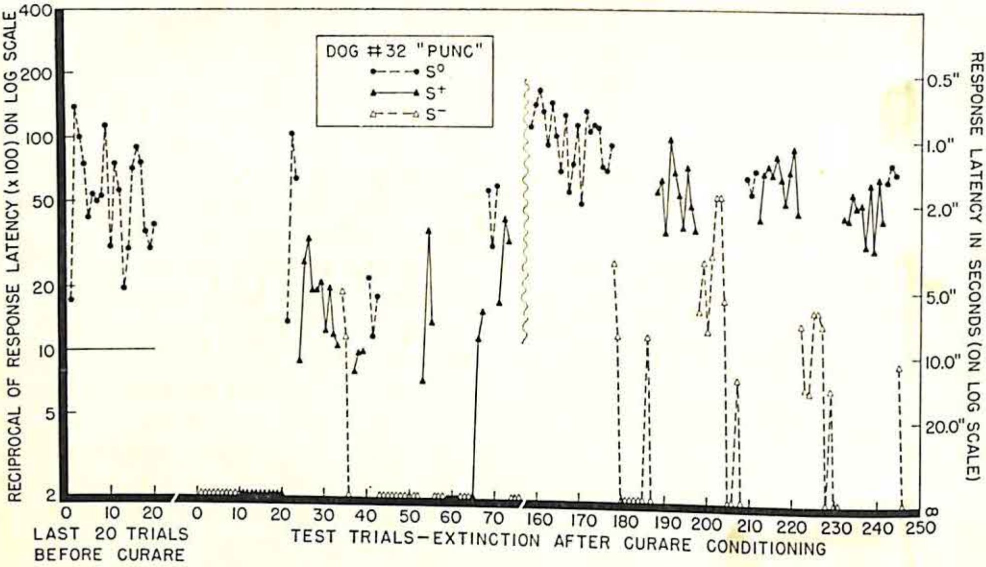


FIG. 5. Transfer effects in Dog #32. Latencies of panel-pressing responses to the original training stimulus, S^o, and to the two stimuli, S^+ and S^- used in the curare-conditioning procedure. (Note that the latencies are expressed in reciprocal form. Note that there is good differentiation of S^+ and S^- only after some tests on S^o are given. Note the long latencies of responses to S^- as contrasted with the quick responses to S^+.)

ditioning Subject #23 was first tested on S^- and failed to respond in 10 tests. Then, when S^+ was next presented, he responded on all 10 tests, but the latencies were longer than those to S^0 had been prior to curare-conditioning. After three failures to respond again to S^-, there were three long-latency responses to S^+ and then there were 10 failures to respond. Extinction on S^+ was not, however, as complete as on S^-. When finally tested on the original S^0, Subject #23 showed stable responding, similar to that shown prior to curare-conditioning. Such tests on S^0 seemed to facilitate subsequent responding to S^+ but not to S^-. There were no responses to S^- during the whole course of testing. However, when S^0 was tested, subsequent tests on S^+ always yielded a few responses followed by extinction. Finally, the response to S^0 extinguished to some extent after S^0 had been presented 45 times over Test Trials 47–99. On Test Trials 108–122 there were 11 failures to respond to S^0. *Thus, even when the panel-pressing avoidance response was in the process of extinguishing, the differentiation between S^- and S^+ was striking.*

Subject #25, shown in Figure 4, like the other subjects, achieved stable avoidance responding prior to curare-conditioning. He showed no sensitization phenomenon prior to curare-conditioning when pretested on S^- and S^+. Subsequent to curare-conditioning, however, there was good evidence of differentiation between S^- and S^+. We started testing with S^- and obtained immediate panel-pressing. The latencies were variable, however, and many of them were longer than those previously seen in response to S^0 prior to the curare-conditioning session. In contrast, responding to S^+ was stable and of the same order of magnitude as that obtained on S^0 prior to curare-conditioning. There were no failures to respond quickly to S^+. But there were, between Test Trials 1–65, 23 failures to respond in 35 presentations of S^-.

We then embarked on a long series of tests, extending over three experimental days, during which we tried to extinguish the panel-pressing response. We alternated long sequences of S^+ and S^0, testing infrequently on S^-, hoping to weaken the panel-pressing response. In Figure 4 we show the results of this testing after 420 trials. The response to S^0 was clearly weaker than it had been. The response to S^- was still absent. In contrast, the response to S^+ was only slightly weaker than it had been about 50 trials earlier. Here, then, was a case where curare-conditioning, even though the panel-pressing response could not be performed during it, produced a stronger avoidance response to S^+ than that to S^0, even though the panel-press responses were made originally to S^0 and not to S^+!

Subject #32 is shown in Figure 5. He, like the other subjects, achieved stable avoidance response latencies prior to curare-conditioning. When tested for sensitization with S^- and S^+, he gave no panel-pressing responses. After curare-conditioning, Subject #32 was first tested with S^-, and he made no panel-pressing responses. Then, when tested with S^+ he again made no responses. Finally, when tested with the original training stimulus, S^0, he responded with characteristic short latencies. Following three quick responses to S^0, he responded with longer latencies to S^+. Then there were two responses to S^-, followed by a failure to respond. From Test Trials 40–76, there were no responses to 16 presentations of S^-; there were 21 responses out of 22 tests on S^+; and

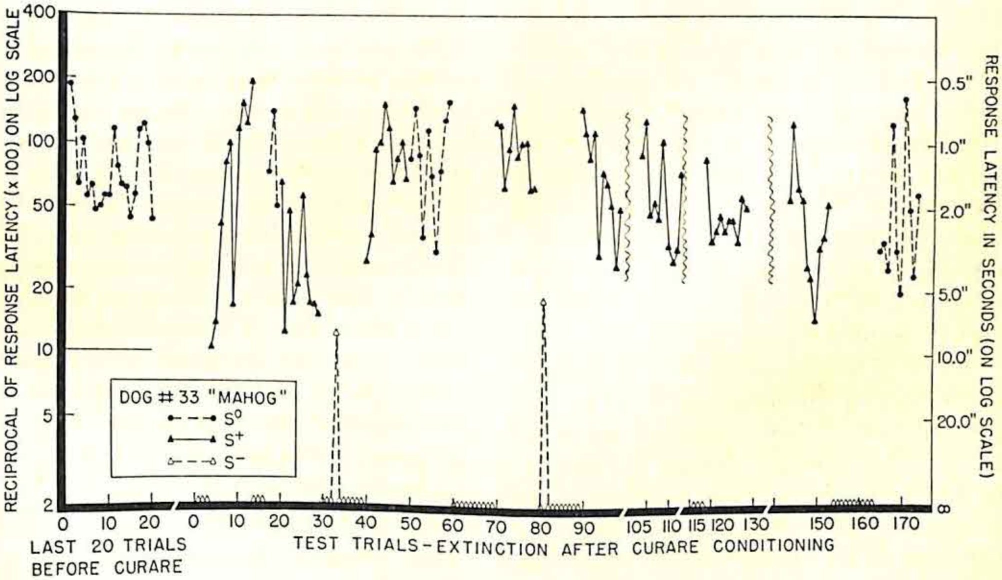


FIG. 6. Transfer effects in Dog #33. Latencies of panel-pressing responses to the original training stimulus, S^0, and to the two stimuli, S^+ and S^-, used in the curare-conditioning procedure. (Note that latencies are expressed in reciprocal form. Note the excellent differentiation of S^+ and S^- at the outset of testing. Note the maintenance of this differentiation over 150 test trials.)

there were six responses out of nine tests on S^0. With repeated testing on all three CSs, there was no apparent extinction trend for S^0 or S^+, while the latencies of response to S^- were either long or infinite. Over Test Trials 158–246, the subject gave 26 short-latency responses to 26 S^0 presentations; he gave 30 short-latency responses to 30 S^+ presentations; and in contrast, he gave 19 responses in 33 presentations of S^-, and these were of long latency. While there was no significant difference between response latencies to S^0 and S^+, the latencies to S^- differed significantly from those of S^+ and S^0. Here, again, there was clear evidence of differentiation in avoidance responding to the two discriminative stimuli used in curare-conditioning.

In Figure 6 are shown data for Subject #33. His avoidance responses prior to curare-conditioning were stable and of short latency. When tested for sensitization by presentations of S^- and S^+, he gave no panel-presses. After the curare-conditioning session, he was first tested on S^-. He failed to respond to three such presentations. Then, when tested with 10 successive presentations of S^+, he responded 10 times; the first responses were sluggish, but they increased in vigor, and the final tests of this sequence yielded latencies of the same magnitude as those characteristic of responding to S^0 prior to curare-conditioning. Then three more S^- presentations elicited no responses. Three S^0 presentations then elicited short-latency panel-presses. Over the course of 173 test stimulus presentations, the responses to S^+ showed little change; the responses to S^0 also remained stable. In contrast, the responses to S^- were practically absent; there were only 2 panel-presses in a total of 49 S^- presentations, and these were sluggish. Here again, there was clear evidence of differentiation between S^+ and S^-.

DISCUSSION

The results show, in all six subjects, that the panel-pressing responses to S^+ were more vigorous, more reliable, less variable, and of shorter latency than those to S^-. Responses to S^+ were also more highly resistent to extinction effects than were those to S^-. On the other hand, the comparison of responses to S^0 and S^+ does not yield a clear conclusion. In some subjects, S^+ was more resistant to extinction effects than was S^0, and in other subjects, the reverse was true. We feel that the number of shocks received in original avoidance training with S^0 may be a factor in producing this variability of outcome. Dogs #32 and #33 received the most shocks in training, and were carried out to a strict avoidance learning criterion, 77 and 88 trials, respectively. These two subjects had the best differentiation between S^- and S^+ in the tests. In addition, the testing procedure following curare-conditioning may be important. Presenting S^0 or S^+, followed by other CSs, seems to elevate the general responding level to all CSs. Presenting S^- followed by the other CSs, seems to depress the general responding level to all CSs. And finally, the 1200-cycle tone seemed slightly better as an S^+ than was the 160-cycle tone. The number of subjects in our experiment is not great enough to prove these points, nor is the design of our experiment adequate for proof. However, the indications are promising enough to pursue further with appropriate experiments.

The theoretical implications of this experiment are clear. Animal subjects can acquire discriminative instrumental response tendencies without the overt responses themselves being exercised in the presence of the discriminative stimuli. The differential strengths of panel-pressing tendencies in the presence of S^- and S^+ could only have been acquired during the discriminative curare-conditioning session. The pretests for sensitization showed no generalization from S^0 either to S^- or to S^+ prior to curare-conditioning. Following curare-conditioning (during which S^+ was paired with shock and S^- was presented without shock pairings), tests on all six subjects showed stronger panel-pressing tendencies to S^+ than to S^-. In some cases responses to S^- were practically absent. Since no panel-pressing could occur during curare-conditioning, there could be no temporal pairing either of S^+ and R_{press} or S^- and R_{press}. The differential, discriminative panel-pressing which revealed itself later must have been due to the Pavlovian procedure of classical, discriminative conditioning. Thus, one might ask, what were the mediators of this special type of transfer of training or thinking? What conditions could have led our dogs to reveal, through their panel-presses, what they had acquired during the conditioning procedure? We feel that a systematic exploration of these questions could lead to a better knowledge of transfer and thought. *By exclusion, we can definitely maintain that differential peripheral skeletal responses (and their associated feedback) could not have been a factor in mediating the discrimination learning which we have clearly demonstrated. Thus, those theories of transfer of training and of thinking which restrict themselves to peripheral skeletal response mediators, and their correlated feedback from the periphery, cannot explain the outcome of our experiment.*

One possible mediation process could reside in ANS conditioning and its correlated feedback from the ANS periphery. There was some evidence, in a few of our subjects, that there was a greater cardiac response to S^+ than to S^- at the end of the curare-conditioning session. (A detailed account of discriminative cardiac conditioning in curarized dogs will appear later in another paper.) Possibly the differential strengths of ANS CRs elicited by S^+ and S^- could serve to mediate the discriminative panel-pressing behavior through correlated visceral sensory feedback processes. If so, we do not know the essential conditions for such mediation. For example, what outcome would be obtained if the subjects were given discriminative curare-conditioning first, then were trained to panel-press to S^0, and then were tested on S^- and S^+ in the normal state? What would have been the outcome if the CS-US interval used in avoidance training to S^0 were shorter than, or longer than, that used in the curare-conditioning session when S^+ was paired with shock? These questions, which definitely suggest a long series of parametric studies, could, if answered, tell us a great deal about conditions for transfer of training and for thinking. For, despite the clear results of this experiment, we still don't have a good explanation for the very "intelligent" behavior of our dogs. What led them to panel-press to S^+? They could have merely howled, struggled in the harness, or have become "frozen" if S^+ had frightened them, and they could have done less of this when S^- was presented. Why did they respond appropriately?

Our subjects clearly "put together" two experiences in such a way that two stimuli, never before associated directly with lever-pressing, elicited differential lever-pressing during the transfer tests. The accomplishment of this transfer phenomenon did not require differential skeletal responding during the differentiation phase. Of course there may be kinds of transfer effects, or types of thinking, for which peripheral skeletal responding during differentiation is necessary or important. The type of phenomenon we have studied is clearly not one of these.

SUMMARY

Dogs were trained to avoid shock in response to a signal. The avoidance response was the pressing of a panel, the signal (S^0) was a light going out, the shock was of 4-milliamperes intensity (applied to the hind toe pads), and the time interval between signal onset and shock onset was 10 seconds. If the correct response occurred during the time interval between signal onset and the usual shock presentation time, no shock was given and the signal was terminated. Intertrial intervals were varied systematically, with a mean of 1.5 minutes.

After the dogs were reliably pressing the panel in response to the signal, with response latencies of 3 seconds or shorter, they were totally paralyzed by curarization. While the dogs were thus completely immobilized under curare, a Pavlovian discriminative conditioning session was carried out. A new signal (S^+) was consistently paired with shock, using a delayed conditioning procedure and a time interval of 10 seconds between S^+ onset and shock onset. The shock duration was 5 seconds. On some trials a contrasting signal (S^-) was presented for 15 seconds, but it was not paired with shock. A sequence of 99 discriminative conditioning trials was presented, ending with an S^+ trial. The S^+ and S^- trials were partly randomized in a special sequence. After this conditioning session, the dogs were given 48 hours in which to recover from the various physiological side effects of curarization. Next they were returned to the training situation in the normal, undrugged state, and the three previously used stimuli (S^0, S^+, and S^-) were presented. The latency of panel-pressing responses was recorded during these tests of the efficacy of the three stimuli.

The dogs responded in a way consistent with their discriminative Pavlovian conditioning experience under curare. There were frequent panel-presses in response to S^0 and S^+, very few in response to S^-. When a dog pressed a panel in the presence of S^-, the latency of the response was often long.

The technique used in this experiment is a special case of training wherein the trainee can do nothing overtly about the information or stimulus sequences as they are presented. It therefore gives the experimenter very precise stimulus control, because the usual overt responses of the subject are not present to modify the stimulus situation in unpredictable ways.

This experiment demonstrates that certain types of transfer of training or problem solving can occur without the benefit of mediation by peripheral skeletal responses or their associated feedback mechanisms. Whether or not peripheral ANS reaction mechanisms play a role in such transfer phenomena remains to be seen, and such a possibility should be seriously explored in future experiments.

REFERENCES

BECK, E. C., & DOTY, R. W. Conditioned flexion reflexes acquired during combined catalepsy and de-efferentation. *J. comp. physiol. Psychol.*, 1957, **50**, 211–216.

BLACK, A. H. The extinction of avoidance responses under curare. *J. comp. physiol. Psychol.*, 1958, **51**, 519–524.

BLACK, A. H., CARLSON, N. J., & SOLOMON, R. L. Exploratory studies of the conditioning of autonomic responses in curarized dogs. *Psychol. Monogr.*, 1962, in press.

CRISLER, C. Salivation is unnecessary for the establishment of the salivary conditioned reflex induced by morphine. *Amer. J. Physiol.*, 1930, **94**, 553–556.

CULLER, E., COAKLEY, J. D., SHURRAGER, P. S., & ADES, H. W. Differential effects of curare upon higher and lower levels of the central nervous system. *Amer. J. Psychol.*, 1939, **52**, 266–273.

ESTES, W. K. Discriminative conditioning: I. A discriminative property of conditioned anticipation. *J. exp. Psychol.*, 1943, **32**, 150–155.

ESTES, W. K. Discriminative conditioning: II. Effects of a Pavlovian conditioned stimulus upon a subsequently established operant response. *J. exp. Psychol.*, 1948, **38**, 173–177.

GIRDEN, E. Cerebral mechanisms in conditioning under curare. *Amer. J. Psychol.*, 1940, **53**, 397–406.

GIRDEN, E. Dissociation of blood pressure conditioned responses under erythroidine. *J. exp. Psychol.*, 1942, **31**, 219–231. (a)

GIRDEN, E. Dissociation of pupillary conditioned responses under erythroidine and curare. *J. exp. Psychol.*, 1942, **31**, 322–332. (b)

GIRDEN, E. Generalized conditioned responses under curare and erythroidine. *J. exp. Psychol.*, 1942, **31**, 105–119. (c)

GIRDEN, E. Effect of striated muscle paralysis induced with erythroidine upon electroencephalogram (E.E.G.). *Proc. Soc. Exp. Biol. Med.*, 1943, **53**, 163–164. (a)

GIRDEN, E. Role of response mechanism in learning and in "excited emotion." *Amer. J. Psychol.*, 1943, **56**, 1–20. (b)

GIRDEN, E. Conditioned responses in curarized monkeys. *Amer. J. Psychol.*, 1947, **60**, 571–587.

GIRDEN, E. The E.E.G. in curarized mammals. *J. Neurophysiol.*, 1948, **11**, 169–173.

GIRDEN, E., & CULLER, E. Conditioned responses in curarized striate muscle in dogs. *J. comp. Psychol.*, 1937, **23**, 261–274.

HARLOW, H. F. The effects of incomplete curare paralysis upon formation and elicitation of conditioned responses in cats. *J. genet. Psychol.*, 1940, **56**, 273–282.

HARLOW, H. F., & SETTLAGE, P. H. The effect of curarization of the fore part of the body upon the retention of conditioned responses in cats. *J. comp. Phychol.*, 1939, **27**, 45–48.

HARLOW, H. F., & STAGNER, R. Effect of complete striate muscle paralysis upon the learning process. *J. exp. Psychol.*, 1933, **16**, 283–294.

LAUER, D. W. The role of the motor response in learning. Unpublished PhD thesis, University of Michigan, 1951. (Microfilm Library, No. 2615)

LIGHT, J. S., & GANTT, W. H. Essential part of reflex arc for establishment of conditioned reflex. Formation of conditioned reflex after exclusion of motor peripheral end. *J. comp. Psychol.*, 1936, **21**, 19–36.

MCINTYRE, A. R., BENNETT, A. L., & HAMILTON, C. Recent advances in the pharmacology of curare. *Ann. N. Y. Acad. Sci.*, 1951, **54**, 301–305.

MORGAN, C. T. The psychophysiology of learning. In S. S. Steven (Ed.), *Handbook of experimental psychology*. New York, Wiley, 1951.

SETTLAGE, P. H. The effect of sodium amytal on the formation and elicitation of conditioned reflexes. *J. comp. Psychol.*, 1936, **22**, 339–343.

SMITH, S. M., BROWN, H. O., TOMAN, J. E. P., & GOODMAN, L. S. The lack of cerebral effects of d-tubocurarine. *Anesthesiology*, 1947, **8**, 1–14.

UNNA, K. R., & PELIKAN, E. W. Evaluation of curarizing drugs in man. VI. Critique of experiments on unanesthetized subjects. *Ann. N. Y. Acad. Sci.*, 1951, **54**, 480–489.

(Received September 21, 1960)

Psychological Review
1962, Vol. 69, No. 3, 220–232

THE ASSOCIATIVE BASIS OF THE CREATIVE PROCESS [1]

SARNOFF A. MEDNICK
University of Michigan

The intent of this paper is the presentation of an associative interpretation of the process of creative thinking. The explanation is not directed to any specific field of application such as art or science but attempts to delineate processes that underlie all creative thought.

The discussion will take the following form. (*a*) First, we will define creative thinking in associative terms and indicate three ways in which creative solutions may be achieved—serendipity, similarity, and mediation. (*b*) This definition will allow us to deduce those individual difference variables which will facilitate creative performance. (*c*) Consideration of the definition of the creative process has suggested an operational statement of the definition in the form of a test. The test will be briefly described along with some preliminary research results. (*d*) The paper will conclude with a discussion of predictions regarding the influence of certain experimentally manipulable variables upon the creative process.

Creative individuals and the processes by which they manifest their creativity have excited a good deal of interest and curiosity. There are extended analyses of novels and novelists, poems and poets, mathematics and mathematicians, both biographical and autobiographical. Perusal of the introspections of manifestly creative individuals uncovers a surprising vein of similarity in the processes they describe (Ghiselin, 1952). Thus we find Albert Einstein's self-searching to suggest that "The psychical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be combined . . . This combinatory play seems to be the essential feature in productive thought." Samuel Taylor Coleridge is described as having developed his ideas in the following manner: "Facts which sank at intervals out of conscious recollection drew together beneath the surface through the almost chemical affinities of common elements." In the field of art, we find André Brèton referring to a collage by Ernst as being distinguished by a "marvelous capacity to grasp two mutually distant realities without going beyond the field of our experience and to draw a spark from the juxtaposition." Most explicit, however, is the oft-quoted statement by the mathematician, Poincaré, who talks about an evening when "ideas rose in crowds; I felt them collide until pairs interlocked so to speak, making a stable combination. By next morning I had established the existence of a class of Fuchsian functions." From these experiences, Poincaré felt that he could state that "to create consists of making new combinations of associative elements which are useful. The mathematical facts worthy of being studied . . . are those which reveal to us unsuspected kinships between other facts well known but wrongly believed to be strangers to one another. Among chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart." An exceptionally compelling illustration of a useful combination of elements "drawn from domains which are far apart" occurs in a line from the poem, "The Monkey Puzzle" by Marianne Moore (1951), "The lion's ferocious chrysanthemum head."

---

[1] The essence of this paper was written while the writer was a Visiting Research Psychologist at the Institute of Personality Assessment and Research, University of California, Berkeley. The author wishes to acknowledge his indebtedness to Benton J. Underwood for his encouragement and the stimulation of his paper, "Orientation to research on thinking" (1952). Work on this material has been supported by the Cooperative Research Program of the Office of Education (Project No. 1073) and the National Science Foundation (Grant G-3855).

We will state our basic hypothesis regarding the nature of creative thinking in the form of a definition. With these introspective statements serving as background, we may proceed to define the creative thinking process as the forming of associative elements into new combinations which either meet specified requirements or are in some way useful. The more mutually remote the elements of the new combination, the more creative the process or solution. An additional criterion of the level of creativeness of a product is described below.

Creative thinking as defined here is distinguished from original thinking by the imposition of requirements on originality. Thus, 7,363,474 is quite an original answer to the problem "How much is 12 + 12?" However, it is only when conditions are such that this answer is useful that we can also call it creative. There are many original ideas expressed in institutions for the mentally ill and mentally retarded; few of these are likely to be creative. There are many fields of creative endeavor in which the usefulness of products would be difficult to measure reliably. While these difficulties must eventually be faced, for the present our research efforts have been concentrated on laboratory situations in which criteria for usefulness can be arbitrarily experimenter-defined and unequivocally explained to the subject. The originality of a response is simply inversely related to its probability in a given population.

It should be pointed out that this definition of creativity is quite similar to basic notions advanced by British associationists from Locke (1690) to Bain (1855), and by those psychologists whose work is based in large measure on their speculations. Freud (1938), Hollingsworth (1928), and Binet (1899) may serve as examples.

WAYS OF ACHIEVING A CREATIVE SOLUTION

In terms of associative theory, we may point to three ways of achieving a creative solution. Generally, any condition or state of the organism which will tend to bring the requisite associative elements into ideational contiguity will increase the probability and speed of a creative solution. Therefore, the following three ways of attaining creative solutions are all methods of bringing the requisite associative elements together.

Serendipity. The requisite associative elements may be evoked contiguously by the contiguous environmental appearance (usually an accidental contiguity) of stimuli which elicit these associative elements. This sort of creative solution is often dubbed serendipitous. This is the manner of discovery to which is popularly attributed such inventions as the X ray and such discoveries as penicillin. One physicist has described how he has reduced serendipity to a method by placing in a fishbowl large numbers of slips of paper, each inscribed with a physical fact. He regularly devotes some time to randomly drawing pairs of these facts from the fishbowl, looking for new and useful combinations. His procedure represents the operational embodiment of this method of achieving creative solutions.

Similarity. The requisite associative elements may be evoked in contiguity as a result of the similarity of the associative elements or the similarity of the stimuli eliciting these associative elements. This mode of creative solution may be encountered in creative writing which exploits homonymity, rhyme, and similarities in the structure and rhythm of words or similarities in the objects which they designate. The contiguous ideational occurrence of such items as alliterative and rhyming associates may be dependent on a factor such as primary stimulus generalization. It seems possible that this means of bringing about contiguity of associational elements may be of considerable importance in those domains of creative effort which are less directly dependent on the manipulation of symbols. Here we might include certain approaches to painting, sculpture, musical composition, and poetry.

Mediation. The requisite associative elements may be evoked in contiguity through the mediation of common elements. This means of bringing the associative elements into contiguity with each other is of great importance in those areas of endeavor where the use of symbols (verbal, mathematical, chemical, etc. . . .) is mandatory. For example, in psychology, the idea of relating reactive inhibition and cortical satiation may have been mediated by the common associates "tiredness" or "fatigue" (Köhler & Fishback, 1950).

INDIVIDUAL DIFFERENCES

From the definition given above, the factors that will make for individual differences in the probability of achieving creative solutions may be deduced. Any ability or tendency which serves to bring otherwise mutually remote ideas into contiguity will facilitate a creative solution; any ability or tendency which serves to keep remote ideas from contiguous evocation will inhibit the creative solution.

Listed below are several illustrative predictions concerning individual differences that one may make from this theoretical orientation.

Need for Associative Elements

It should be clear that an individual without the requisite elements in his response repertoire will not be able to combine them so as to arrive at a creative solution. An architect who does not know of the existence of a new material can hardly be expected to use it creatively.

Associative Hierarchy

The organization of an individual's associations will influence the probability and speed of attainment of a creative solution. There is a whole family of predictions that one may draw from this concept of the associative hierarchy. As an initial example, let us take the question of the manner in which the associative strength around ideas is distributed. If we present an individual with the word "table," what sort of associative responses does he make? The individual who tends to be restricted to the stereotyped responses, such as "chair," may be characterized as having an associative hierarchy with a steep slope (see Figure 1). That is, when you get past the first one or two conventional responses to the stimulus, the individual's associative strengths to other words or ideas (lower in the hierarchy) drops rapidly. We can also conceive of a second sort of individual whose associative hierarchy is characterized by a rather flat slope. This is an individual who perhaps also has as his strongest response the conventional chair. But for him this response is not overly dominant and so it is more likely that he will be able to get to the less probable, more remote kinds of associations to table. It is among these more remote responses that the requisite elements and mediating terms for a creative solution will be lurking. This slope factor may be related to the mathematical analysis of associative production developed by Bousfield, Sedgewick, and Cohen (1954). It probably is closely approximated by their constant, *m*, measuring rate of depletion of the associative reservoir. They found a high negative correlation between rate of association and total number of associations. It would be predicted from Figure 1 that the high creative subject (flat hierarchy) would respond relatively slowly and steadily and emit many responses while the low creative subject (steep hierarchy) would respond at a higher rate but emit fewer responses.

It would be predicted that the greater the concentration of associative strength in a small number of stereotyped associative responses (steep hierarchy) the less probable it is that the individual will attain the creative solution. Thus, the word association behavior of the high creative individual should be characterized by less stereotypy and commonality. This last prediction is supported by a study by Mednick, Gough, and Woodworth (Mednick, 1958). Research scientists rated for creativity were divided into relatively high ($N = 15$) and relatively low ($N = 15$) groups. The low creatives gave more stereotyped responses on 80% of a group of 36 test words from the Kent-Rosanoff list. (These test words were chosen for their tendency to elicit stereotyped responses. Stereotypy was defined by the Minnesota Kent-Rosanoff Word Association Norms, Russell & Jenkins, 1954). It should be pointed out that these results lend themselves to another possible interpretation. The highly creative individual may also have a steep hierarchy but a deviant one. That is, his most dominant associative response may be quite strong but quite different from the popular, dominant associative response. There are different predictions that can be made for the flat-associative-hierarchy creative and the steep-deviant associative-hierarchy creative. The latter is more likely to be the one-shot producer (a not uncommon phenomenon among novelists). If he does create further products, they will tend to resemble closely the first product. The former is more likely to be a multiproducer; he is more likely to produce in a variety of avenues of creative expression.

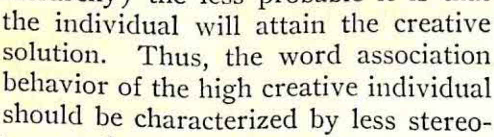


FIG. 1. Associative hierarchies around the word "table."

The prediction suggesting an expectation of less creativity from an individual with a high concentration of associative strength in a few responses leads to another prediction. The greater the number of instances in which an individual has solved problems with given materials in a certain manner, the less is the likelihood of his attaining a creative solution using these materials. Such an individual will "know the meaning" of the elements of the subject matter. That is, he will have a steep associative hierarchy around these elements. An example of the operation of this principle recently occurred to the writer while teaching an honors freshman introductory course in psychology at the University of Michigan. I was giving a well known interpretation of a well known experiment in stimulus generalization when interrupted by a student who calmly stated that the interpretation was in error. After a few minutes of blustering I asked him to explain. His explanation proved him to be correct. I had been dealing with this material for years and "knew" the "correct" interpretation; for him this material was new, he had a low, flat associative hierarchy. Thus, if a newcomer to a field has the requisite information, he is more likely to achieve a creative solution than a long-time worker in the field. This may be the reason that theoretical physicists and master chess players are often said to have passed their prime by the age of 25.

Number of Associations

The greater the number of associations that an individual has to the requisite elements of a problem, the greater the probability of his reaching a creative solution. This variable is not independent of the preceding one since an individual with a high concentration of associative strength in few associative responses is not likely to have a proliferation of associations. The more associates which are evoked by a requisite element of a problem, the more likely it is that an associate will exist which will serve as a mediating bridge to another requisite element, facilitating combination. It seems likely that this variable will not be related to speed of creative solution since it may take a good deal of time to get to the mediating links.

Cognitive or Personality Styles

Previously learned or innately predisposed methods of approaching problems will influence the probability of a creative solution. If the requisite associational elements of a new and useful combination are probable associates of the concrete representations of relevant aspects of the problem, an individual with a predominately "perceptual" approach will be more likely to reach a creative solution. If, however, the requisite associational elements are not elicited as responses to these concrete representations or if there is no concrete representation then an individual with a "conceptual" approach will be more likely to reach a creative solution.

Another cognitive style of importance may lie along the "visualizer-verbalizer" dimension. The visualizer is one who tends to call up relatively complete memorial sensory representations of the relevant concrete aspects of problems. If the problem deals with horses, he tends to picture a horse in terms of its sensory qualities. On the other hand, the verbalizer explores the problem by associating with words around the word "horse." If the requisite elements are high in his verbal associative hierarchy to the word horse, the verbalizer will be more likely to attain a creative solution; the visualizer may be thrown off or at least delayed by many false leads. On the other hand, if a requisite verbal associative response to the word horse is very low, or not present in the verbalizer's hierarchy, then the visualizer will be more likely to attain the creative solution. It is therefore clear that some types of problems will be solved more easily by the visualizer and some by the verbalizer.

Factors such as these (admittedly very poorly defined) may be partly responsible for differential aptitudes for creative work in differing fields.

Selection of the Creative Combination

The creative combination of elements is only one among the many which may present themselves to the subject. How or why is the creative combination selected? Some speculations regarding this problem follow. The explanation of the process of selection may be related to the nature of the problem. Problems either entail a specific and relatively objective set of testable criteria (Paint a realistic portrait of this individual. Design a refrigerator so that it will be automatically free of frost.) or they do not (The chemist mixes two liquids out of curiosity. The painter dabs hopefully at a fresh canvas waiting for an idea. The psychologist tosses a new test into a correlation matrix). When specific criteria are provided, they form an important part of the stimulus set which is determining which associative elements are being elicited and thus becoming eligible for entering into combination with other elements. Important sets of associations to each of these combinations are the consequents of the combinations. The set of consequents for each combination (If I put x, y, and z together, a and b will happen) is continually compared with the set of requirements of the problem. When the set of consequents of a new combination achieves a close fit with the set of problem requirements, this combination is selected. When there is complete overlap of sets, "search behavior" is terminated. As with the other requisite elements of the problem, individual differences in this case will vary with (among other things) the structure of the associational hierarchies around the requirements of the problem. When the refrigerator-defroster problem was presented to an undergraduate class almost all of the proferred solutions were based on the principle of ridding the refrigerator of already heavily accumulated frost. A couple of individuals (possibly familiar with the defroster principles presently in use) suggested methods which disposed of the frost before it built up to an overly annoying level. In addition to these there were two unique responses, i.e., a "new" method of preventing moisture from condensing in the freezer compartment, and a method of allowing frost to accumulate but limiting the location of accumulation to a small box which could be regularly and conveniently removed and emptied. Thus it may be seen that an individual's associations to the requirements may be characterized as to their stereotypy; the imposed requirements of the problem may be viewed as part of the requisite elements in the situation. The earlier theoretical statements concerning these elements may be seen as being relevant here. The foregoing suggests an explanation of the selection process for the case where the subject must hunt for a combination of elements which will satisfy given criteria. In the case where no criteria are specified, the subject is typically producing random combinations of elements; the task of selection in this case consists in finding relevant criteria for the given partial products.

If we may continue along a bit further with this example of the defroster, we may begin to see some glimmerings of a solution to the most serious problem in research on creative thinking—how may we determine to what degree behavior is creative? We have suggested one criterion in our hypothesis. In the following an additional criterion is developed. To begin with let us examine the requirements as originally stated—"Design a refrigerator so that it is automatically free of frost." The first thing that strikes us is that while some requirements have been stated, there are even more that are strongly implied and essential, many that are desirable, and a number that we would only become aware of after some method of satisfying them had been suggested.

Let us examine some possible solutions:

1. Simply refraining from opening the refrigerator door would solve the problem as stated since this would prevent moisture from entering and condensing as frost. This solution meets many of the implied requirements. It is cheap, convenient, effective, does not require special training, etc. . . . However, it is not an optimal solution since it violates one essential, implied requirement—the usefulness of the refrigerator must not be impaired. (This is the cutting-off-your-nose-to-spite-your-face solution.)

2. A primitive solution is the hammer-and-screwdriver method. This is tried and true and meets many of the essential requirements. It falls down in that it is inconvenient, messy, uneconomical (when caked with frost, the refrigerator unit is very inefficient), endangers the mechanism, and is hardly automatic.

3. In a refrigerator we once owned another solution was used. The opening and closing of the refrigerator door operated a counter. At a certain count the refrigerator unit was automatically heated and the melted water evaporated outside the refrigerator. The superiority of this solution is immediately apparent. The source of this superiority lies in the number of requirements which it meets. It is economical, automatic, convenient, peculiarly appropriate (the operation of the heating element is contingent upon the number of door openings. The amount of frost accumulated is also in part dependent on the number of door openings.), does not interfere with the normal use of the refrigerator, and does not require special training. Note that the principle behind this highly creative solution (not allowing massive build-ups of frost) was infrequently suggested in the classroom group. However, this solution is not wholly successful at meeting some criteria. The frequent heating and cooling may injure frozen food stored near the heating element. Secondly, since the heating process must be brief and mild, it is inevitable that not all frost is removed. While this solution does effectively curtail the number of defrostings, it does not eliminate them completely. It is clear that a method which would encompass all of the advantages of the "counter" method, but which would, in addition, eliminate defrosting altogether would be even more creative. What is suggested by this discussion is that the creativeness of a product is some function of the number of requirements that the product meets. The most ready application of this definition will be in laboratory research in which tasks, solutions, and requirements may be arbitrarily constructed and varied.

A TEST OF CREATIVITY

The definition of the creative process has suggested a way of testing for individual differences in creativity. The test items are intended to require the testee to perform creatively. That is, he is asked to form associative elements into new combinations by providing mediating connective links. Since the test situation is contrived, the combination must meet specified criteria that are experimenter imposed.

The definition dictates the structure of the test. We must provide stimulus items from two mutually distant realities and ask the subject to "draw a spark from their juxtaposition." To state it more usefully, we must provide stimulus elements from mutually remote associative clusters and have the subject find a criteria-meeting mediating link which combines them. A first problem concerns the type of material of which the stimulus item should be composed. If the test is to be appropriate for all fields of creative endeavor, the material must either be nonsensical so as to avoid bias favoring any specific means of creative expression, or it must be so common in society that familiarity could be assumed to be high across fields of interest. The problems involved in constructing the nonsense materials so as to avoid favoring any interest groups soon proved to be apparently insurmountable. This left us searching for materials with which most individuals in the culture could claim acquaintance; this, in turn, brought us to verbal materials.

While it may be true that certain occupational groups have extensive experience in dealing with words, there are some verbal associative habits that could reasonably be assumed to be familiar to almost all individuals that have been brought up in this (USA) culture. Among such habits are the associative bonds between words like "ham and eggs," "bed-bug," "pool-hall," "hound-dog," "whole-wheat," "chorus-girl," "kill-joy," and "red-hot." These became the materials for the test.

Having decided on the materials, the test almost constructed itself in accordance with the definition. Several words from mutually distant associative clusters must be presented to the subject; his task must be to provide mediating links between them. Further, (a factor of extreme importance), the mediating link must be strictly associative rather than being of a sort that follows elaborate rules of logic, concept formation, or problem solving. In their final (or at least present) form, the test items consist of sets of three words drawn from mutally remote associative cluster. One example might be:

Example 1:	rat	blue	cottage

The subject is required to find a fourth word which could serve as a specific kind of associative connective link between these disparate words. The answer to Example 1 is "cheese." "Cheese" is a word which is present in the word pairs "rat-cheese," "blue-cheese," and "cottage-cheese." The subject is presented with several examples so that he has an adequate opportunity to achieve the specific set necessary for the task.

Example 2:	railroad	girl	class
Example 3:	surprise	line	birthday
Example 4:	wheel	electric	high
Example 5:	out	dog	cat[2]

(None of these examples is a test item from any form of the actual test.) The two college level forms of the test (one coauthored by Sharon Halpern and the other by Martha T. Mednick) have 30 items each; the subject is allowed 40 minutes; his score is the number right.

The test, called the Remote Associates Test (RAT), has some interesting correlations with other measures.

Comparisons with Criteria. A study was conducted at the College of Architecture, University of California, Berkeley, by the writer and Sharon Halpern. Ratings of creativity by faculty members of the College who taught the Design courses were correlated with RAT scores. These ratings form an unusually excellent criterion of creative performance since the raters had been advising and evaluating the students in the creation of new designs and models of structures. They had been working with these students for at least a year and in many cases two or more. The ratings and RAT scores correlated significantly ($r = .70$, $df = 19$, $p < .01$) In this study an early form of the RAT was used.

---

[2] Answers to sample RAT items: 2. working; 3. party; 4. chair or wire; 5. house.

The RAT was administered to a group of first year psychology graduate students at the University of Michigan whose native language was American English ($N = 35$). Faculty research supervisors (who had been directing the independent research efforts of the students), rated the eight highest and eight lowest RAT scorers either "high" or "low" in research creativity (no middle category allowed). Research creativity was defined as being demonstrated if the student developed new research methods and/or pulled together disparate theory or research areas in useful and original ways. Of the 16 research supervisors, one felt that he had not had enough contact with his student to make the judgment. His student was a low RAT scorer. Of the eight high RAT scorers, six were rated high on research creativity and two were rated low; of the seven low RAT scorers, only one was rated high, the other six being rated low. By Fisher's exact test the probability of these events occurring by chance is less than .05. Miller Analogies Test (MAT) scores were available for these students. Of the seven high MAT scorers, three were rated high on research creativity; of the eight low MAT scorers, four were rated high in research creativity.

Reliability. The Spearman-Brown reliability of the RAT was .92 in one sample (289 women, almost all the students at an Eastern women's college, tested as part of a project under the direction of Theodore Newcomb) and .91 in another (215 men tested at the University of Michigan as part of a project under the direction of Warren T. Norman).

Correlation with Grades. One of the present forms of the RAT correlated negatively with the first-two-year grade point averages of a group of undergraduates at a large Eastern technology college. ($r = -.27$, $N = 74$, $p < .05$). This same correlation was obtained with the summer grades of a smaller group, $N = 34$, of summer students at a large Eastern liberal arts college (not statistically significant in this case). In a study by Miller (1960) it was found that high RAT scorers tended to get higher grades from teachers rated as flexible than from teachers rated as dogmatic. Low RAT scorers received higher grades from teachers rated dogmatic than from teachers rated as flexible.

Correlation with Social Attitudes and Occupational Interests. It is clear that creative individuals must have access to improbable associative responses. Kowalski (1960) hypothesized that this is a general tendency which also manifests itself in their attitudes and interests. She tested and interviewed 15 high RAT scoring and 15 low RAT scoring undergraduate women. The two groups had radically differing views on sexual morality and women's rights. The views of the high creatives were more atypical and "liberal" ($U = 37$, $p < .001$). On the Strong Vocational Interest Blank, Mens' Form (SVIB), the high creative group showed "significantly higher interest on the artist ($p < .05$), psychologist ($p < .005$), physician ($p < .025$), mathematician ($p < .025$), and author-journalist ($p < .05$) keys. The low creative group showed higher interest on the farmer ($p < .05$), math-physical science high school teacher ($p < .05$), office man ($p < .05$), and pharmacist ($p < .01$) keys" (p. 19). (These are the probability values of obtained chi squares.) The only one of these keys related to ACE scores was that of physician. "The commonality of these interest patterns was evaluated by noting the per cent overlap of the specific key with the general population of interest expression. For example, the Farmer key overlaps 45% with the general population, while the Artist key overlaps 20%" (p. 20). The significant keys of the higher creatives had significantly less commonality than the significant keys of the low creatives. These differences were independent of the influence of intelligence as measured by the ACE.

Associative Behavior. In the discussion of illustrative predictions it was suggested that highly creative individuals would be characterized by a flat associative hierarchy rather than a steep associative hierarchy. Further, it was proposed that the greater the number of associations that an individual has to the requisite elements of a problem, the greater the probability of his reaching a creative solution. From these two independent statements it may be deduced that when required to display his reservoir of associations to single stimulus words, the highly creative individual will have greater access to less probable associates and therefore produce a greater number of associates. A study by Craig and Manis (1960 unpublished[3]) supports this deduction. Thirty-eight college students had the RAT and an associative task administered to them. In the associative task they were given 1 minute to write as many associates as they could to each of 20 words. The correlation of the number of such associates with RAT scores was .38 ($p < .01$).

In two related studies, Karp (1960) and Kowalski (1960) found RAT scores to be directly related to the originality and quantity of anagrams constructed using the test word "Generation." In the Karp study 40 undergraduates were given 5 minutes to produce as many four letter anagrams from the test word as they could. The productions were scored for quantity (number of acceptable answers) and originality (a weighted score for each response was developed from the frequency with which the response was given by the 40 subjects). The correlation of the RAT with the quantity scores was .44 ($p < .01$); the correlation of the RAT with the originality score was .37 ($p < .05$). Kowalski presented the anagrams task to 15 high RAT scorers and to 15 low RAT scorers, giving them 5 minutes to produce words of any length from the test word "Generation." In this study originality was measured by computing the percentage of responses given by an individual which had not been given by any other of the 30 subjects. The difference on this measure between high and low RAT scorers was significant ($U = 68$, $p < .04$). "Only four subjects in the low creative group gave *any* original responses at all while eleven subjects in the high creative group did" (p. 19).

At the Institute of Personality Assessment and Research, University of California, Berkeley, the RAT was included as part of the assessment battery administered to a group of 40 highly eminent architects. The RAT correlated .31 with the Originality (O-I) Scale of the IPAR Questionnaire Scale and −.31 with the total Conformity Score obtained in the Crutchfield Conformity Experiment (Crutchfield, 1955). Interviewers rated high scorers as significantly higher in "graceful and well-coordinated in movement" and "reticent and taciturn in speech." The college grade point average which the subjects reported correlated −.34 with

---

[3] Craig, M., & Manis, M. Prediction of scores on the Remote Associates Test by size of response repertoire. Unpublished manuscript, 1960.

RAT score, a result which tends to confirm findings reported above.

EXPERIMENTALLY MANIPULABLE VARIABLES

While only one experimental study (described below) which makes use of this general framework has been completed in this laboratory, it may be useful to indicate briefly the kinds of experimental investigation it suggests.

Massed vs. Distributed Work Sessions. Total time of work being equal, massed sessions of creative work should be more successful than distributed sessions. There are two reasons why this would be so. The first is that the individual making use of the massed session technique is more likely to achieve temporal contiguity of the requisite associative elements within a single intensive work period than is an individual who has distributed his work in shorter periods over several days. Secondly, it may take some time for an individual to work on a problem enough to go beyond its obvious aspects. In the first hour of work, he may get through only the conventional and stereotyped associations to the elements of the problem, while it is perhaps in the later stages of intensive work on a problem that one can begin to entertain the more remote associations that are evoked by elements of the problem. It is, of course, among these remote associations that the key to the creative solution will lie.

Warmup. In creative work a warmup session should serve to arouse the more remote associations to the requisite elements of the problem. While their work has gone considerably beyond the problem of warmup, Maltzman, Bogartz, and Breger (1958) have demonstrated that the repeated elicitation of different word associations to the same stimulus words does indeed tend to produce remote associations to these stimulus words. Further, this induced originality tends to transfer to other relatively unrelated tasks presented after this associative warmup. Associative warmup of this type should become more effective as the warmup stimuli are more similar to the task materials. It may be that the effects of warmup will prove to be a further advantage that massed sessions have over distributed sessions for creative productivity.

Stereotyping Associative Responses. As stated above, if an individual's associative response to a stimulus element of a creative problem is of excessive strength, this will tend to reduce the likelihood of occurrence of more remote associative responses. This will reduce the probability and speed of creative solution. It would therefore be predicted that extensive training of a specific response other than a requisite one to a stimulus element of a problem requiring a creative solution should retard later attempts at solution of the problem. This prediction is related to the concept of "functional fixedness" introduced by Duncker (1945). Birch and Rabinowitz (1951) and Adamson and Taylor (1954) completed experiments which are related to this prediction. Their test situation was the two string problem. The subjects are asked to tie together two strings suspended from the ceiling. When the subject grasps one string he finds that the other string is hanging out of his reach. The solution to the problem requires the subject to attach a weight to one of the strings, get the weight swinging and catch it while holding the other string. Various objects can be used as a weight. The subjects that had been pretrained by having them use a switch in its usual manner of functioning tended to be unlikely to use it as a weight. They had developed strong response strength for the association "switch-close circuit" which had reduced the probability of the remote association "switch-weight."

Another feasible experimental approach would make use of the RAT item as a creative task and test the influence upon it of certain variables. For example, the words of which an item is composed may be presented at varying rates to test the massed trials hypothesis. In addition, various pretraining conditions may be evaluated in terms of their effectiveness in increasing or decreasing RAT performance.

Another possible experimental approach would entail separating out high and low RAT scorers and observing the differential effect of certain variables upon their behavior. In an experiment just completed Houston and Mednick (in press) postulated that an important motive impelling the behavior of the creative individual was a need for improbable associative stimulation. It was reasoned that if such stimulation were supplied, it would tend to satisfy this need and be reinforcing. Further, if such stimulation regularly followed a given response the high creative individual should tend to learn that response. Thirty high and 30 low RAT scorers were asked to read aloud only one of two typed words on a 3 × 5 card. Excepting buffer items and including 40 pairs aimed at gauging the free operant level of noun-choice, there were 160 pairs of words, each pair consisting of a noun and a nonnoun (verb, adjective, adverb, etc.). If a subject in the experimental group (15 high RAT subjects, 15 low RAT subjects) responded with the noun member of a pair, the experimenter responded with an improbable association; if the subject chose the nonnoun, the experimenter responded with the most probable association. In the control group (15 high RAT, 15 low RAT) both the nouns and the nonnouns were invariably followed by their most probable associate. Associative probabilities were obtained from the Russell and Jenkins (1954) and Deese (1960) norms. If the improbable response was satisfying a need, the probability of noun-choice should increase over the 160 trials. It did significantly in the high RAT experimental group; the low RAT experimental group showed a decrease. The high and low RAT control groups showed no reliable change.

Some of the positions which have been taken in this paper are assumptions and not deductions. As more data are gathered some of these assumptions will assume the status of facts, some will be revised. For example, the opening paragraph suggests that the paper is not meant to apply only to one field of creative endeavor but attempts to delineate processes that underlie all creative thought. This may require modification. The explanation may fit the process of scientific discovery and not be appropriate to discussions of painting or music. For the present (paradoxically enough), the more encompassing assumptions seem more parsimonious. It may eventually turn out (as is hinted at in the body of the paper) that the differences between the fields are more determined by differences in suitability of the three means of achieving contiguity, i.e., serendipity, similarity, and mediation.

SUMMARY

An associative theory of creative thinking has been outlined. Differences between high creatives and low creatives have been predicted along specified dimensions. Predictions have been made regarding the effect on the creative process of some experimentally manipulable variables.

The associative definition of the creative process has taken the operational

form of a test. Some preliminary research with this test is described.

REFERENCES

ADAMSON, R. E., & TAYLOR, D. W. Functional fixedness as related to elapsed time and to set. *J. exp. Psychol.,* 1954, **47**, 122–126.

BAIN, A. *The senses and the intellect.* 1855.

BINET, A. *The psychology of reasoning.* Chicago: Open Court, 1899.

BIRCH, H. G., & RABINOWITZ, H. S. The negative effect of previous experience on productive thinking. *J. exp. Psychol.,* 1951, **41**, 121–125.

BOUSFIELD, W. A., SEDGEWICK, C. H. W., & COHEN, B. H. Certain temporal characteristics of the recall of verbal associates. *Amer. J. Psychol.,* 1954, **57**, 111–118.

CRUTCHFIELD, R. Conformity and character. *Amer. Psychologist,* 1955, **10**, 191–198.

DEESE, J. Word association norms. Author, 1960. (Ditto)

DUNCKER, K. On problem-solving. *Psychol. Monogr.* 1945, **58**(5, Whole No. 270).

FREUD, S. Wit and its relation to the unconscious. In, *The basic writings of Sigmund Freud.* New York: Modern Library, 1938.

GHISELIN, B. *The creative process.* Berkeley: Univer. California Press, 1952.

HOLLINGSWORTH, H. L. *Psychology: Its facts and principles.* New York: Appleton, 1928.

HOUSTON, J. P. & MEDNICK, S. A. Creativity and the need for novelty. *J. abnorm. soc. Psychol.,* in press.

KARP, S. A validity study of a measure of creativity. Senior honors thesis, University of Michigan, 1960.

KÖHLER, W., & FISHBACK, J. The destruction of the Müller-Lyer illusion in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.,* 1950, **40**, 398–410.

KOWALSKI, J. Attitudes and occupational interests of creative individuals. Senior honor thesis, University of Michigan, 1960.

LOCKE, J. *Essay concerning the human understanding.* 1690.

MALTZMAN, I., BOGARTZ, W., & BREGER, L. A procedure for increasing word association originality and its transfer effects. *J. exp. Psychol.,* 1958, **56**, 392–398.

MEDNICK, S. A. An orientation to research in creativity. (Res. Memo. No. 2) Berkeley, Calif.: University of California, Institute of Personality Assessment and Research, 1958.

MILLER, B. A study of creativity in college students and teaching method types. Senior honor thesis, University of Michigan, 1960.

MOORE, MARIANNE. The monkey puzzle. In, *Collected poems of Marianne Moore.* New York: Macmillan, 1951.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr 66216.

UNDERWOOD, B. J. An orientation to research on thinking. *Psychol. Rev.,* 1952, **59**, 209–220.

(Received September 22, 1960)

Psychological Review
1962, Vol. 69, No. 3, 233–238

REANALYSIS OF "MEANINGFULNESS AND VERBAL LEARNING"

RONALD C. JOHNSON

University of Hawaii

Letter combinations such as nonsense syllables and trigrams (three letter combinations that may or may not form words or syllables) have a number of characteristics found to influence rate of learning. Those characteristics that have been investigated most thoroughly may all be referred to as frequency variables. These variables are:

1. Frequency, per se: A number of possible measures of frequency exist. Some trigrams form words. For these trigrams, Thorndike-Lorge (1944) frequencies can be obtained. Some trigrams and nonsense syllables form sounds or syllables that occur in the English language, and it is possible to obtain some measure of the frequency of this occurrence. The HAV-LAV differentiation of nonsense syllables according to whether or not they occur as the first three letters of English words is essentially a frequency measure. It has proven useful in learning experiments in which nonsense syllables are used as stimuli (e.g., Lindley, 1960) and might also serve as a trigram frequency measure. Finally, the frequency of a trigram or nonsense syllable might be defined as the frequency that the three letters of any given stimulus occur in contiguity in English words.

2. Association value (a): The a of a stimulus is generally established by determining the proportion of subjects who have an association (even if they cannot tell what this association is) to a stimulus within a rather short (2–4-second) time interval. Association value does not directly measure the number of associations evoked by a stimulus, although syllables that evoke *an* association in a high proportion of subjects generally evoke *more* associations, if presented for a longer time interval, than do syllables for which few subjects have any associations.

3. Meaningfulness (m): Noble (1952) has proposed as a measure of m the mean number of associations (excluding clang and certain other categories of association) that a verbal stimulus can evoke from subjects in a fixed time interval. While a is established by determining whether a stimulus can evoke an association, m is arrived at from the mean number of these associations.

All of these frequency variables influence rate of verbal learning. All of these variables are substantially related to one another. The frequency of occurrence of nonsense syllables as three letter sequences in words is related to the a of these syllables (Underwood, 1959), and the Thorndike-Lorge frequency of words is closely related to m (Johnson, Frincke, & Martin, unpublished[1]; Underwood, 1959). Association value (a) and m are also related (Noble, Stockwell, & Pryor, 1957). While these three measures are related, it seems dangerous to assume that because of this relation they are synonymous. Aside from nonequivalent operations used in obtaining these measures, we have ample precedent in studies of the relation of m and familiarity

---

[1] Johnson, R. C., Frincke, G., & Martin, Lea. Meaningfulness, frequency, and affective character of words as related to response measures. Unpublished manuscript, 1960.

(Noble, 1953) and of *m* and *a* (Noble, Stockwell, & Pryor, 1957) for suspecting it to be in error to lump these measures together as equivalent or synonymous.

The most ambitious recent study of verbal learning is probably that of Underwood and Schulz (1960) as described in their book, *Meaningfulness and Verbal Learning*. A number of the experiments reported by Underwood and Schulz involve the learning of lists of trigrams by various groups of subjects. For Underwood and Schulz, a trigram consists of any three successive letters in a word. For example, the word "learning" contains the following trigrams: LEA, EAR, ARN, RNI, NIN, ING. Trigrams were obtained from a large sample of words. Trigram frequency measures were obtained, with frequency being defined as the number of times that the various three letter combinations occurred in contiguity in the sample of words from which they were obtained. The term meaning or M is used interchangeably to denote meaningfulness (*m*) and association value (*a*). While frequency and M are discussed separately by Underwood and Schulz, the assumption that they appear to make is that the relation between these two (or three) variables is so close that separate analyses of the data regarding these variables are unnecessary. Pronunciability ratings were obtained from samples of subjects for those trigrams used in a number of learning experiments.

The trigram experiments center around the problem of determining whether frequency-M or pronunciability can account for the majority of variance in rate of verbal learning. In the first of Underwood and Schulz's trigram learning experiments (Experiment 6, p. 144ff.), four lists of eight trigrams each were learned by various samples of subjects. Taking the lists as a whole, frequency was significantly related to rate of learning, but pronunciability showed an even closer relation. Three lists of 12 trigrams each served as stimuli in Experiment 11 (p. 169ff.). No significant relation was obtained between trigram frequency and rate of learning. The relation between trigram pronunciability and rate of learning was highly significant. Underwood and Schulz say, "Clearly, we have no alternative but to conclude that for these lists the relation between frequency and learning is essentially zero" (p. 171). Other experiments yielded similar results and Underwood and Schulz state that M is unrelated to rate of learning. They say:

> Rather, we believe that the apparent casual status of M derives from a certain amount of covariation with pronunciability, and if this covariation is removed, only the relationship between pronunciability and learning will hold up (p. 192).

(This removal of covariance does indeed reduce the relation of frequency (and M) to approximately zero—e.g., see p. 194. The question, in partial correlation, is what variable to partial out.) They again refer to pronunciability, saying:

> It appears then, quite unexpectedly and quite without theoretical direction, that we have stumbled upon an attribute of verbal units which has more predictive power by far than any other attribute we have discovered or methodically set about to measure. (p. 197)

Further investigations, using generated trigrams, confirmed the belief that of the different variables studied, pronunciability was the most closely related to rate of learning. Underwood and Schulz's rejection of frequency, meaningfulness, and association value as major variables in verbal learning is indeed astounding since this rejection runs counter to the previous interpretations of a most extensive body of data.

It is, therefore, something of a disappointment to find that Underwood and Schulz once more accept frequency as the key variable in verbal learning on p. 289ff. They say that if one defines trigram frequency as the frequency that the three-letter combination occurs as a sound in the English language, rather than using frequency of contiguity as a measure of frequency, then frequency may serve as an adequate prediction of rate of verbal learning.

As pointed out in this paper, there are a number of possible measures of trigram frequency. One does not have to be too gestalt oriented to have considered the possibility that the frequency with which trigrams appear as units of English speech is a better measure of frequency, in terms of predictive power, than is frequency of contiguity of three-letter sequences. While Underwood and Schulz, in their analysis, define frequency in terms of contiguity, this writer will define frequency as the frequency of occurrence of trigrams as discrete or relatively discrete sound units in spoken English. The purpose of this paper is to reanalyze some of Underwood and Schulz's data from this point of view. Experiments 6 and 11 seem representative of their trigram learning experiments and shall form the basis of the major portion of this reanalysis.

EXPERIMENT 1

One syllable words form relatively discrete sound units. Sounded syllables in words form somewhat less discrete units, while three letter combinations that are neither words nor sounded syllables are not sound units of English speech. It is relatively simple to select the three letter words from trigram lists used in Experiments 6 and 11. The one exception to this simplicity in selection is the trigram "QUE," used in Experiment 6, which does not appear in Webster's *New World Dictionary of the American Language,* but does appear among the 30,000 most frequent words in the Thorndike-Lorge tables. The writer, decided, on the basis of its relatively high frequency, to count QUE as a word. The next problem was to separate syllables from nonsyllables. The purpose of the first experiment reported in this paper was to separate those trigrams that occur as syllables (sound units that do not, in themselves, form words) from those that do not occur as sound units, and to determine whether there are differences in the rate that words, syllables, and nonsyllables are learned.

Method

Subjects. The subjects were 51 upper division psychology students.

Materials. The materials were those 52 trigrams from Underwood and Schulz's Experiments 6 and 11 that do not form English words.

Procedure. The subjects were presented with the trigrams in a double column on a single dittoed page in the same order that these trigrams appeared in Underwood and Schulz's book. They were told:

> Look at the first of these three letter combinations. Say it to yourself. Does this combination occur as a sounded syllable in English words? If it does, put a check after it. Do this for all three letter combinations on the page.

Results and Discussion

Whether due to perversity or to misunderstanding, such improbable combinations as CQU were described as syllables by several subjects. For this reason it was decided to count three-letter combinations as syllables appearing in the English language only if rated as being syllables by at least five subjects. The words, syllables, and nonsyllables from Experiments 6 and 11 are as follows:

EXPERIMENT 6

Words—WAS, BUT, MAN, WHO, ART, STY, VIZ, QUE
Syllables—ZON, CYR, STI, JOK, EIG, SOU, VER, YAL, CHA, ING
Nonsyllables—RCH, XPE, IFO, TLY, DGM, XPE, GHT, ULD, CKB, MPT, NDF, XPO, MBK, DFL

EXPERIMENT 11

Words—BOY, URN, HER, HAT, CUT, ITS, PAR, LED
Syllables—PLO, VIF, ING, EST, WHE, SOU, FRO, UND, VER, CHI, ENT, JUM, ROP, ISH
Nonsyllables—NDE, OUS, WHA, ATI, CQU, DGM, YIN, ALI, OMP, MPA, YLV, ABL, NCE, TIO

Kruskal and Wallis *H* tests show that the words, syllables, and nonsyllables used in Experiment 6 differ significantly in pronunciability ($H = 21.7$, $df = 2$, $p < .001$) and in mean correct anticipations ($H = 15.3$, $df = 2$, $p < .001$). The words, syllables, and nonsyllables in Experiment 11 differ significantly in pronunciability ($H = 24.2$, $df = 2$, $p < .001$) and in mean correct anticipations ($H = 13.2$, $df = 2$, $p < .005$).

If we divide the words, syllables, and nonsyllables of each list into these three categories and divide the trigrams in each category into those above and below the median in pronunciability for their group, we obtain the data presented in Figure 1.

It is apparent that Underwood and Schulz, in obtaining pronunciability ratings of these trigrams, ranged words, syllables, and nonsyllables along a continuum, while the experimenter, using another technique, divided them into three discrete groups. The significance figures show that the two measures are substantially related. The method used in differentiating between these trigrams would depend on the definition and the measure of trigram frequency.

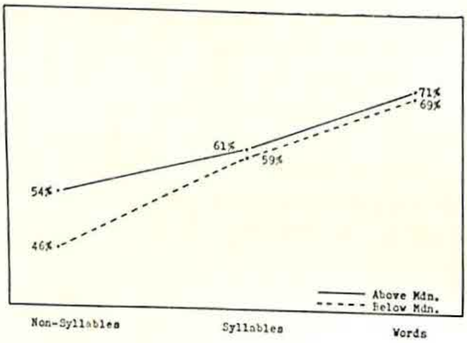


FIG. 1. Mean percentage of correct anticipations for nonsyllables, syllables, and words above and below median in pronunciability.

It seems clear, from this graphic representation as well as from the *H* tests, that trigrams that form words are learned more rapidly than trigrams that form syllables (a somewhat less discrete unit of speech)—contrary to Underwood and Schulz's belief as stated in p. 284.—and that trigrams that form syllables are learned more rapidly than are trigrams judged by subjects as not forming syllables in the English language. When the data are analyzed in this way, pronunciability no longer appears as a major variable.

Whether frequency or pronunciability is the predictor of rate of learning is a result of the definition of frequency used. If frequency is defined as the number of contiguous occurrences of three-letter combinations, then pronunciability seems to be the significant variable. If frequency is defined as the number of occurrences of three-letter combinations at different levels of discreteness in the English language, then frequency seems to be the most significant variable while pronunciability merely serves as an indicator as to whether a trigram is a word, syllable, or nonsyllable.

EXPERIMENT 2

Underwood and Schulz, on p. 280, state that the frequency (of contiguity) hypothesis broke down most completely in multiple letter units in the broad range between the extremes of well integrated and poorly integrated units. Most of those trigrams rated by the subjects as appearing as syllables in the English language fall into this middle range of integration. For this reason, the experimenter investigated further the characteristics of those trigrams used in Experiments 6 and 11 that were rated as being syllables. The experimenter, using the same subjects that had rated the trigrams of Experiments 6 and 11 and using the same procedure, also determined which of the trigrams used in Experiment 15 (p. 245ff.) formed syllables. Five or more subjects rated each of the following trigrams used in Experiment 15 as occurring as syllables in the English language: ING, CED, JAD, ITE, SUL, NUW, MEL, DOK, NOP, SOM.

Method

Subjects. The subjects were 15 introductory psychology students.

Materials. The trigrams from Experiments 6, 11, and 15 that formed syllables but did not form words were placed on dittoed sheets, in random order, 10 trigrams to a page.

Procedure. The sheets were passed out to the subjects, face down. The subjects were cautioned to leave the sheets face down till they were told to turn them over. The experimenter then told them:

> What I want you to do is help me in a learning experiment. You have a list of three-letter combinations that form syllables. I am going to give you 20 seconds for each syllable. In that 20 seconds I want you to say the syllable to yourself in as many ways as you wish. Then I want you to write down, following the syllable, all the words you can think of that contain the sound of the syllable. For example, if the first syllable was L-O-G, then you would say this three-letter combination to yourself and then write down the words that contain L-O-G, as it is sounded, for example, log, logic psychology, logey, and logger. [This example was put on the blackboard.] You are to do this for each syllable. When I say "start" turn your paper over and begin with the first syllable. When I say "stop" stop writing. When I say "start" again, begin with the second syllable. Go through the entire list in this way. Do not look ahead.

The subjects signified that they understood the directions and were administered the list. They were timed by stop watch. Their responses were scored.

Result and Discussion

The mean number of words produced for each of the syllables is presented in Table 1.

The rank-order correlation between the number of words produced for each syllable and the mean number of correct anticipations made by Underwood and Schulz's subjects for the same syllable were +.48 for the syllables used in Experiment 6; +.64 for the syllables used in Experiment 11, and +.75 for the syllables used in Experiment 15. The last two of these correlations are significant beyond the .05 level of confidence. The responses to this produc-

TABLE 1

MEAN WORDS PRODUCED FOR TRIGRAM SYLLABLES

Experiment 6		Experiment 11		Experiment 15	
Syllable	Mean produced words	Syllable	Mean produced words	Syllable	Mean produced words
ZON	1.27	PLO	1.93	ING	4.93
CYR	0.87	VIF	0.87	CED	2.00
STI	2.53	ING	4.93	JAD	1.00
JOK	2.33	EST	3.67	ITE	2.80
EIG	1.47	WHE	2.87	SUL	2.33
SOU	2.27	SOU	2.27	NUW	0.40
VER	2.53	FRO	2.60	MEL	2.47
YAL	1.47	UND	3.00	DOK	3.13
CHA	2.67	VER	2.53	NOP	0.73
ING	4.93	CHI	3.13	SOM	2.87
		ENT	4.33		
		JUM	2.53		
		ROP	2.07		
		ISH	3.40		

tion technique seems somewhat different than, yet associated with, measures of association value, frequency, and meaning. The number of words that the subjects can produce that contain a syllable seems closely associated with the rate at which the syllable is learned.

CONCLUSION

Underwood and Schulz, defining frequency as the frequency that trigrams appear contiguously in English words, found pronunciability to be significantly related to rate of learning, while frequency was not a major variable. This experimenter defined frequency as the frequency that a trigram appeared as a discrete or relatively discrete sound unit in spoken English. Using the approach taken in this paper, frequency is most significantly related to rate of verbal learning while pronunciability is not a major variable. While it may be that other attributes of trigrams (such as *m* or *a*) which covary with the frequency that trigrams occur as sounds in the English language are more basic in influencing rate of verbal learning, one can say, with certainty, that frequency, as defined herein, is an adequate predictor of rate of verbal learning.

REFERENCES

LINDLEY, R. Association value and familiarity in serial verbal learning. *J. exp. Psychol.*, 1960, **59**, 366–370.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.

NOBLE, C. E., STOCKWELL, F. E., & PRYOR, M. W. Meaningfulness (*m'*) and association value (*a*) in paired associate syllable learning. *Psychol. Rep.*, 1957, **3**, 441–452.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Columbia Univer. Press, 1944.

UNDERWOOD, B. J. Verbal learning and the educative process. *Harvard educ. Rev.*, 1959, **29**, 107–117.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received October 14, 1960)

Psychological Review
1962, Vol. 69, No. 3, 239–240

THEORETICAL NOTES

A RESPONSE TO STRICKER'S COMMENTS ON "WORD VALUES, WORD FREQUENCY, AND VISUAL DURATION THRESHOLDS"

RONALD C. JOHNSON
University of Hawaii
CALVIN W. THOMSON
San Jose State College
AND GERALD FRINCKE
University of Illinois

As Stricker (1961) notes, throughout our study we (Johnson, Thomson, & Frincke, 1960) equate word value with the rating of the word on the good-bad scale of the semantic differential. He says, further, that:

> word *value* has been used in the literature to represent a dimension different from that tapped by the *evaluative* scale of the semantic differential, and while the two may share some commonality, it does not seem entirely appropriate to use the latter in order to operationally define the former.

Stricker then demonstrates that word value, as defined in previous visual duration threshold studies, refers to a motivational (e.g., Allport-Vernon) and not an attitudinal (e.g., semantic differential) dimension, and that motivational and attitudinal dimensions of value(?)or evaluation(?) or "X" are probably not synonymous. He says that while we demonstrate that there is a relation between word goodness and word frequency (and also between both goodness and frequency and threshold), we do not demonstrate a relation between word frequency and word *value* (as Stricker prefers to define value).

We cannot deny that the operational definition of "value" varied between our study and that used in previous studies conducted in this area. As we see it, science consists, to a large degree, of figuring out ways of testing the previously untestable. If operationally defining black as white aids the scientist in this endeavor, then it is quite legitimate to do so, *so long as you make your definitions clear*. While the word "value" has been used in the past to refer to other dimensions than those investigated in our study, we do not believe that we, or others, are bound to traditional usage, so long as we define our terms. We plead guilty to departing from prior usage of the word "value," but do not believe this to be reprehensible.

We believe that our shift in operational definition of "value" is defensible on several grounds. The first is that so long as value is defined in terms of such dimensions as Allport-Vernon scores, then it appears to have been clearly demonstrated (Postman & Schneider, 1951; Solomon & Howes, 1951) that the two variables of word frequency and word value are so inextricably intertwined that experimental tests of the influence of frequency and of value on threshold yield results that are difficult, if not impossible, to interpret in any unequivocal sense. By redefining value, we believe that we have made the value-frequency problem more easily testable than was previously the case. If we have, then a redefinition seems legitimate. If Stricker prefers to use value in another, more traditional sense, it is his privilege to do so, and his responsibility to design an experiment testing the influence of value, as he defines it, on frequency and on threshold.

A second defense of our shift in operational definition of value also comes from the ground rules of science. It is generally believed that a major aim of science is to stimulate further scientific endeavor, thus adding to our knowledge. It seems likely that many experimenters were discouraged from entering the value-frequency controversy, since after the Postman and Schneider and the Solomon and Howes studies cited above it became fairly apparent that determining the influence of value on threshold was difficult, if not impossible, so long as value was defined in terms of Allport-Vernon scores. As a result, little experimentation was done in this area, and for a number of years the frequency explanation was (quite correctly, on the basis of the available evidence) generally accepted as being the most parsimonious one available in dealing with the results of visual duration threshold experiments. Recent experiments (Johnson, Frincke, & Martin, 1961; Newbigging, 1961a, 1961b) stimulated, in part, from the study and from the definition of value with which Stricker finds fault, suggest that frequency is not a significant variable, while value, as we define value, is significant (although there is some question as to *why* it is significant) in determining threshold.

We cannot believe that any redefinition of word value that serves to increase the testability of certain experimental questions, generates research, and provides us with a different and, we believe, more correct view of reality, is to be rejected merely because it violates traditional usage.

REFERENCES

JOHNSON, R. C., FRINCKE, G., & MARTIN, LEA. Meaningfulness, frequency, and affective character of words as related to visual duration threshold. *Canad. J. Psychol.*, 1961, **15**, 199–204.

JOHNSON, R. C., THOMSON, C. W., & FRINCKE, G. Word values, word frequency, and visual duration thresholds. *Psychol. Rev.*, 1960, **67**, 332–342.

NEWBIGGING, P. L. The perceptual redintegration of frequent and infrequent words. *Canad. J. Psychol.*, 1961, **15**, 123–132. (a)

NEWBIGGING, P. L. The perceptual redintegration of words which differ in connotative meaning. *Canad. J. Psychol.*, 1961, **15**, 133–142. (b)

POSTMAN, L., & SCHNEIDER, B. H. Personal value, visual recognition, and recall. *Psychol Rev.*, 1951, **58**, 271–284.

SOLOMON, R. L., & HOWES, D. H. Word frequency, word value, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.

STRICKER, G. Word values, word frequency, and visual duration thresholds: A comment. *Psychol. Rev.*, 1961, **68**, 420–422.

(Early publication received January 9, 1962)

Psychological Review
1962, Vol. 69, No. 3, 241-245

AN APPLICATION OF LEVINE'S MODEL FOR HYPOTHESIS BEHAVIOR TO SERIAL REVERSAL LEARNING[1]

BENJAMIN H. PUBOLS, JR.

University of Wisconsin[2]

Recently, Levine (1959a, 1959b) described a mathematical model for hypothesis behavior in discrimination learning-set experiments. By defining a hypothesis (H) as "a specifiable pattern of response to a selected stimulus set" (1959b, p. 353), and making certain assumptions regarding their nature and conditions of occurrence, Levine was able to evaluate the strengths of various hypotheses (e.g., position preference, stimulus preference, problem solution) throughout learning-set formation, to analyze learning-set curves in terms of the contributions of component Hs, and to synthesize learning-set surves given H strengths. In his dissertation (1959a, p. 58), Levine suggested that this approach might also be appropriate to a description of the finding of North (1950), Pubols (1957), and others, that rats show progressive improvement in serial position reversal learning.

The published model is not directly applicable to serial reversal learning data, but since its appearance Levine[3] has developed a general model for hypothesis behavior, applicable to both learning-set and serial reversal experiments. A recent experiment by the writer (Pubols, 1962) yielded data in terms of which this model might be tested.

In this experiment, three groups of 10 rats each were given 410 trials in a one-unit alley **Y** maze. Members of Group 10 received 10 trials per reversal for original learning and 40 reversals of a position habit; members of Group 20 received 20 trials per reversal for original learning and 20 reversals; and members of Group 40 received 40 trials per reversal for original learning and 10 reversals (in all cases, only 10 trials were given on the final reversal, as this sufficed for measurement purposes). Empirical results of the study have been reported elsewhere (Pubols, 1962). Here, we will be concerned with the application of Levine's model to the serial reversal data. But first, a brief summary of the model seems in order.

SUMMARY OF LEVINE'S GENERAL MODEL

Levine's model is restricted to the operation of H behavior over the first three trials of a problem. He makes the following assumptions: (*a*) if an H occurs on a problem, it will be manifest over all three trials of that problem; (*b*) while the determiners of response on **Trial n may** include stimuli associated with Trial $n-1$ (as in single alternation), the response will be independent of Trial $n-2$ stimuli (thus excluding from consideration an H such as double alternation); (*c*) all Hs are mutually exclusive; (*d*) the sum of the probabilities of all operative Hs is one; (*e*) the selection and definition of Hs for a given application is independent of the characteristics of the model; and (*f*) "a given H has the same strength no matter in which cell [sequence] it appears" (Levine, 1959b, p. 358).

The operations involved in estimating H strengths differ somewhat for learning-set and for serial reversal data, and hence these will be considered separately. For any three-trial problem in the learning-set situation, there are four reward sequences, four response sequences, and two conditions of Trial 1 outcome (rewarded or nonrewarded). This yields 32 possible sequences, or cells. The procedure for evaluating H strengths basically involves two steps. The first step utilizes a simplifying assumption, which may be called the assumption of symmetry: in a learning-set experiment, each problem involves a new pair of discriminanda, and hence it is reasonable to assume that, on each problem, the probability of a Trial 1 positive sequence equals the probability of a Trial 1 negative sequence:

---

[1] This investigation was supported by PHS Grant M-4046(A), National Institute of Mental Health, United States Public Health Service. The author would like to express his appreciation to Marvin Levine and Alfred Etaugh for critical readings of preliminary drafts of this paper.

[2] Laboratory of Neurophysiology, Department of Physiology.

[3] Personal communication, January 11, 1961.

$$P(+_1)=P(-_1)$$

This being the case, the first step involves the calculation of conditional probabilities of empirical sequences, the probability of the empirical outcome on Trials 2 and 3, given the reward sequence and outcome on Trial 1. Operationally, the frequency of occurrence of a given sequence is divided by the frequency of occurrence of all sequences having the same reward sequence and Trial 1 outcome. The second step then involves the simultaneous solution of N equations, representing the N hypotheses whose strengths are being evaluated, taking into account differences in the number of cells associated with each H. The final result, then, is a statement of the probability of each H in each block of problems analyzed.

Turning now to serial reversal data, the total number of sequences reduces from 32 to 8. This follows from the fact that there is only one reward sequence involved; that is, on a given reversal, reward is always on the same side or position. Further, response sequence and reward outcome become completely confounded. Defined in terms of reward outcomes, the eight serial reversal sequences are, for Trial 1 negative: $---$, $--+$, $-+-$, and $-++$; for Trial 1 positive: $+++$, $++-$, $+-+$, and $+--$.

In the case of serial reversal (and also, incidentally, oddity learning) experiments, the assumption of symmetry is not met. As serial reversal learning progresses, the probability of a Trial 1 negative sequence increases while that of a Trial 1 positive sequence decreases; hence, an H will come to have a greater strength in the former case. When the assumption of symmetry is not met, Levine[4] has shown that the use of conditional probabilities of empirical outcomes is not appropriate; rather absolute probabilities should be calculated.

Thus, for the evaluation of serial reversal data, absolute probabilities of sequence occurrence should be utilized. The absolute probability of a given sequence is defined as the ratio of the frequency of that sequence to the frequency of all sequences:

$$P(i\text{th sequence}) = \frac{\text{frequency of } i\text{th sequence}}{\text{frequency of all sequences}} \quad [1]$$

APPLICATION OF THE GENERAL MODEL

The analysis of H behavior for the data of the Pubols experiment is based on reversals only, not on the original learning. Reversals were grouped into blocks of five, yielding 50 scores per block (5 reversals × 10 rats). This gives, for Groups 10, 20, and 40, respectively, eight, four, and two blocks for analysis. Table 1 outlines the hypotheses chosen for analysis, along with other relevant information. Equation 1 now becomes:

$$P(i\text{th sequence}) = \frac{\text{frequency of } i\text{th sequence}}{50} \quad [2]$$

After these sequence probabilities are computed, r, a, b, p_2, and p_3 are obtained in the usual way, and the final H probabilities are defined as $R=8r$, $A=2a$, and $B=2b$, p_2 and p_3 remaining unchanged.

Results of the application of the model are shown in Figures 1 and 2. Figure 1

---

[4] Personal communication, January 11, 1961.

TABLE 1

THE HYPOTHESES EVALUATED IN THE PRESENT ANALYSIS

Hypothesis	Probability symbol	Associated sequences	Number of sequences
Random responding	R	All sequences	8
Position preference	A	$---$, $+++$	2
Position alternation	B	$-+-$, $+-+$	2
Trial 2 problem solution	p_2	$-++$	1
Trial 3 problem solution	p_3	$--+$	1

presents the experimental data as analyzed into the various Hs. Inspection of the figure reveals an interesting happening. Whereas all three groups show interreversal learning (progressive increase in $p_2 + p_3$), this learning is correlated, in different groups, with the diminution of different "incorrect" Hs. Thus, for Group 10, interreversal learning is accompanied by a systematic decrease in position alternation, *B*; for Group 20, there is a decrease primarily in random responding, *R*; and for Group 40, it is position preference, *A*, which systematically decreases.

In order to obtain some indication of the reliability of the trends just described, each curve of Figure 1 was subjected to an individual trend test.[5] For these tests, H strengths in each rat on each block of reversals were computed, and the error term was the blocks by subjects interaction. As expected, in Group 10, the trend in the problem solution curve was significant ($p < .01$), as was the trend in the position alternation curve ($p < .001$). For Group 20, only

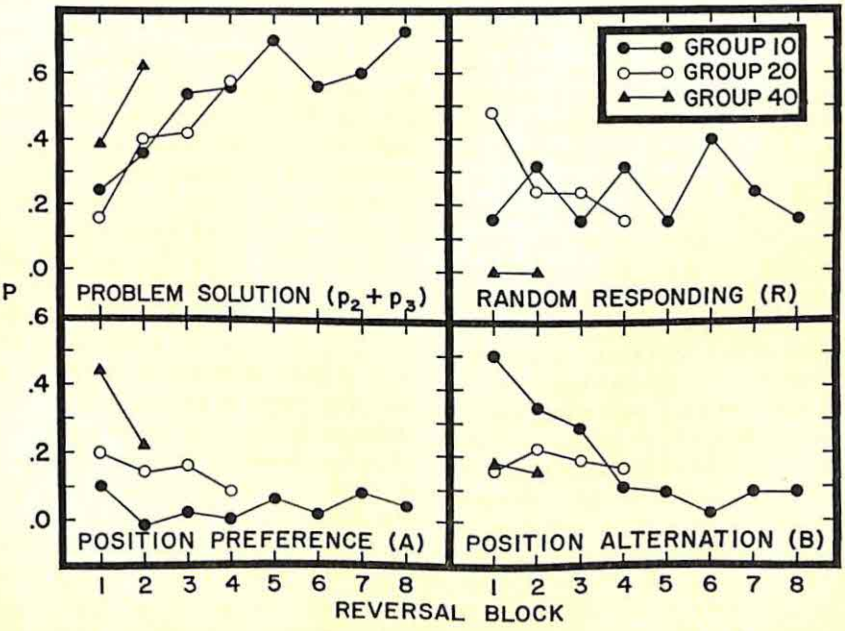


FIG. 1. Hypothesis probabilities as a function of reversal block.

---

[5] Because of different numbers of reversal blocks in the three groups, and because within an individual subject all hypotheses are not statistically independent, it was not possible to run any over-all trend tests.

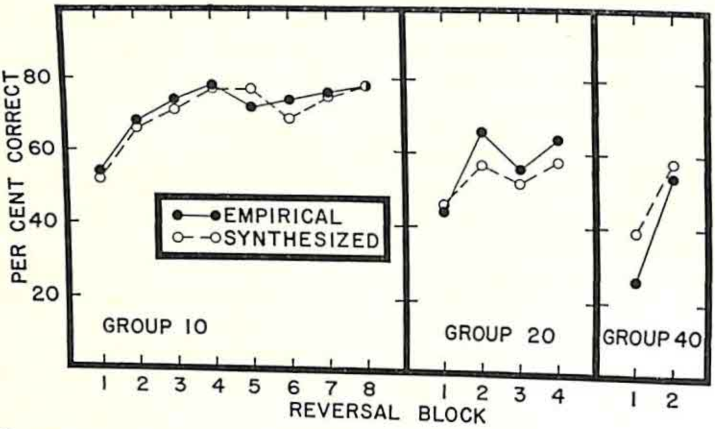


FIG. 2. Comparison of empirical and synthesized serial reversal functions, in terms of Trial 2 percentage correct.

the trend in the problem solution curve was significant ($p < .05$), and for Group 40, only the trend in the position preference curve was significant ($p < .05$). The failure of significance in the case of the Group 20 random responding, and the Group 40 problem solution curves can probably be attributed to the loss in degrees of freedom resulting from the combination of reversals into blocks. Certainly, in the case of Group 40, there is independent evidence (Pubols, 1962) that interreversal learning was genuine.

It seems that, as the number of trials per reversal increases, the initially dominant hypothesis changes from position alternation, possibly to random responding, and then to position preference. The strengths of these Hs were, in the indicated groups, high from the very first reversal, suggesting the importance of the number of trials on the original problem. In the case of Group 40, the position preference hypothesis manifests itself exclusively in the sequence, $- - -$, which indicates strong perseveration of the formerly correct response. With fewer trials per reversal, this perseverative effect is less strong and incorrect responding is manifest in an alternation or a random pattern.

Figure 2 compares conventional (empirical) serial reversal curves based on Trial 2 performance with curves synthesized according to Levine's Equation 6 (1959b, p. 363). The fit is seen to be best for Group 10, poorest for Group 40. Trial 3 curves were also synthesized, but the fit was less close in all three cases, primarily by way of overestimation.

DISCUSSION

The general model appears to have as much utility when applied to serial reversal learning as when applied to learning-set formation. Where the probability of a Trial 1 negative sequence equals the probability of a Trial 1 positive sequence, the strength of Hs may be evaluated in terms of either absolute or conditional probabilities of empirical outcomes. In contrast, where Trial 1 positive and negative sequences have unequal probabilities, the strength of Hs is appropriately evaluated only in terms of absolute probabilities.

There yet remains one discrepancy between the model and serial reversal data. This discrepancy involves the assumption that a given H has the same strength in all cells in which it appears. This assumption implies, for example, that $P(- - -) = P(+ + +)$, and that $P(- + -) = P(+ - +)$. That these equalities do not hold follows from the fact that $P(-_1) > P(+_1)$. However, the analytical utility of the model, as indicated by Figure 1, and the close correspondence between empirical and synthesized functions, as indicated by Figure 2, suggest that, in practice, this may not be a serious shortcoming.

Levine (1959b, pp. 365–366) compares explanations of "learning to learn" that have been suggested in the literature. Harlow's view (1959; see also Harlow & Hicks, 1957) is that interproblem (and presumably interreversal) learning involves the weakening and eventual elimination of incorrect response tendencies, or, in the present context, Hs not consistently reinforced. An opposing view (Restle, 1958) emphasizes the strengthening of rewarded responses (Hs). While our data show that increases in p_2 and p_3 are accompanied by decreases in the strength of other Hs, it cannot be concluded, as Harlow implies, that the latter is the cause of the former. Nor, of course, can the reverse be concluded. We are in agreement with Levine (1959b, pp. 365–366) that, in terms of the model for hypothesis behavior, learning to learn is best described as the strengthening of certain Hs through 100% reinforcement and the weakening (extinction) of other Hs through 50% reinforcement.

REFERENCES

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

HARLOW, H. F., & HICKS, L. H. Discrimination learning theory: Uniprocess vs. duoprocess. *Psychol. Rev.*, 1957, **64**, 104–109.

LEVINE, M. A model of hypothesis behavior in discrimination learning set. Unpublished PhD dissertation, University of Wisconsin, 1959. (a)

LEVINE, M. A model of hypothesis behavior in discrimination learning set. *Psychol. Rev.*, 1959, **66**, 353–366. (b)

NORTH, A. J. Improvement in successive discrimination reversals. *J. comp. physiol. Psychol.*, 1950, **43**, 442–460.

PUBOLS, B. H. Successive discrimination reversal learning in the white rat: A comparison of two procedures. *J. comp. physiol. Psychol.*, 1957, **50**, 319–322.

PUBOLS, B. H. Serial reversal learning as a function of the number of trials per reversal. *J. comp. physiol. Psychol.*, 1962, **55**, 66–68.

RESTLE, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, **65**, 77–91.

(Received February 23, 1961)

Psychological Review
1962, Vol. 69, No. 3, 246–256

A COMPARISON OF THE POWER OF THE *U* AND *t* TESTS[1]

C. ALAN BONEAU
Duke University

In a recent paper (Boneau, 1960), the author summarized the results of a number of theoretical and empirical studies dealing with the effects of violations of assumptions underlying the *t* test. It was concluded that the *t* test is remarkably unaffected by the two common violations: sampling from populations having unequal variances and sampling from nonnormal distributions. One who uses the *t* test can be reasonably sure that the probability of rejecting a true null hypothesis is close to the alpha value he selects for his experiment even though he may have misgivings about the assumptions upon which the *t* test is based. As a result of these considerations, a recommendation was made in the previous paper to the effect that even when the assumptions are not met (except under special conditions) the *t* test and the *F* test on means of the analysis of variance be used without those attendant feelings of turpitude which can be attributed to an introjection of the strictures of the proponents of nonparametric methods.

This recommendation was based upon an assumption that the *t* and *F* tests, because they make effective use of the information in the sample and have other desirable properties, should be more powerful techniques than nonparametric competitors. (For readable explanations and discussions of the power of a test see among others Siegel, 1956 or Walker & Lev, 1953.) That is to say, if the null hypothesis is false—if there are true differences between means—the *t* test should signal the detection of small differences by yielding significant results more frequently than should comparable nonparametric methods.

Theoretically, the *t* test is more powerful than any of the usually utilized tests when the assumptions underlying it are met. It is also true, however, that when sampling is from certain nonnormal distributions, other tests may be more powerful than the *t* test. The Wilcoxon-Mann-Whitney *U* test (Wilcoxon, 1945; Mann & Whitney, 1947) for example, in one pathological case, theoretically is infinitely more powerful than the *t* test. Theoretically also, the power of the *U* test is never less than .83 of that of the *t* test (Hodges & Lehman, 1956). In fact even in the case for which the *t* test is designed (normality and equal variances) the *U* test by one measure of the relative power of the two tests is 95% as powerful as the *t* test (Hodges & Lehman, 1956).

Such theoretical statements about relative power of tests are based upon mathematical limiting processes involving conditions which are not representative of most practical situations, for example, infinitely large sample sizes and arbitrarily small differences between population means: (Dixon, 1954; Hodges & Lehman, 1956; and Mood, 1954). Statements as to the relative efficiency in general of various nonparametric competitors of the *t* test are scattered throughout the literature (Dixon, 1954; Hodges & Lehman, 1956; Lehman, 1953; Mood, 1954, to mention only the relatively accessible ones). They seem not, however, to have permeated effectively that hard core of statistical lore which the research psychologist musters in an attempt to wrest truth from chaff.

The present paper is intended to present the facts (culled from the literature as well as manufactured for the purpose) about the power of the *t* test and, in particular, how that power compares with the power of a specific nonparametric competitor in various practical situations. This presentation is meant to temper in part the implications of the previous paper (Boneau, 1960) that the *t* test should be used whenever possible. It would seem that here, as in other areas of human endeavor, a little discretion may pay off.

---

[1] This project was supported by NSF Grant G-9592 to the author and by NSF Grant G-6694 to the Duke University Digital Computing Laboratory. The author wishes to express appreciation to Thomas M. Gallie, Director of the Laboratory, for his cooperation and assistance.

Attention will be focused upon the *U* test, a worthy protagonist whose principal strengths vis-à-vis the *t* test have already been mentioned. The *U* test, or equivalent versions of procedures based upon ranked scores, has been invented several times in the history of statistics, first by Deuchler (1914), but later by Wilcoxon (1945), Mann and Whitney (1947), among others. (See Kruskal, 1957, for historical discussion.) As used in the present context, the statistic *U* is computed by determining the number of scores in the second sample which are exceeded by each score in the first sample. The sum of all such counts summed over the scores in the first sample is called *U*, tables for which have been developed by Mann and Whitney (1947), extended by Auble (1953), and made readily accessible by Siegel (1956). Wilcoxon's *T* test (1945), although limited to equal-size samples, gives exactly equivalent results even though it is computed in a different fashion.

The null distribution of *U* may be derived from the assumption that ranks are assigned to the two samples on a random basis such that every combination of ranks among samples is equally likely. For example, given that n_1, the first sample size, is 2 and n_2 is 3, the possible values of the two ranks in the first sample (assuming no ties) are 1 and 2, 1 and 3, 1 and 4, 1 and 5, 2 and 3, 2 and 4, 2 and 5, 3 and 4, 3 and 5, and finally, 4 and 5. By the definition above, these lead to *U* values of 6, 5, 4, 3, 4, 3, 2, 2, 1, and 0, respectively. If all of these combinations are equally likely—the two samples came from the same distribution for example—the probability of getting values as extreme as 6 or 0 by chance is the sum of the individual probabilities:

$$1/10 + 1/10 = 1/5$$

By the logic of statistical decisions, however, one attributes extreme values of *U* not to these accidents of random sampling which occur with known probability if the null hypothesis is true, but to actual differences between the distributions. In actuality, differences in distribution usually lead to non-equally-likely combinations of ranks. Ranks 1 and 2 occur together with relatively greater frequency as the difference between the means of the populations increases and as a result more frequent extreme *U* values occur. For this reason *U* is generally considered to be a test of displacement or shift of distributions, the main focus being on differences in central tendency. Note also that discrepancies of variance tend also to produce non-equally-like combinations of ranks. If two populations have the same mean, the values 1 and 5 tend more to occur together as the size of the variance of the first increases relative to the size of that of the second. However, these combinations give rise to middling values of *U* and hence are ignored, in effect, by the *U* test. We shall see, however, that there are cases for which differences in distribution other than central tendency affect the value of *U*.

One further statement might be made about the sensitivity of the *U* test in specific cases. Since essentially it is based only upon ranks, first and last scores get rank 1 and *n* whether they are close to the mean or several standard deviations away. Thus the occasional score which is apparently not in the distribution but which furnishes no real justification for exclusion is treated as a member of the ingroup by the *U* test but as the pariah it may well be by the *t*.

The method of the present paper is quite similar to the approach followed by the author in the previous paper dealing with the probability of rejecting the null hypothesis if it is true. In that study populations having specified characteristics were constructed and the values of *t* arising from the differences between means of random samples drawn from them were computed. The empirical probability of rejecting the null hypothesis was obtained by determining the proportion of sample t values falling outside the ordinary tabled values for the appropriate number of degrees of freedom (i.e., falling in the region of rejection). Since the null hypothesis was indeed true, these empirical probabilities or proportions could be compared with the nominal values to determine the effects of modifying the specified characteristics of the populations in such a way as to violate the assumptions underlying the t test. In the present study, the concern is with the proportion of obtained t's and Us falling in the region of rejection (or critical region) when the null hypothesis is false—when there is a built-in, specified difference between means. To generate this information, the only required addition to the previous t test program is a provision for changing the mean of the first sample to any value desired.

The program using the IBM 650 Electronic Data Process System for generating t's from random deviates was discussed in detail in the previous paper (Boneau, 1960). To summarize briefly, 10-digit random numbers were generated by a multiplication process. These were converted into random deviates from a specified population by a table-look-up procedure; the random deviates were injected into the computing formula for the t test for the difference between means of two independent samples; and the resulting t value was sorted and tallied on an internally contained table within the computer.

The program designed for the present study for the U test utilized the existing random number and random deviate generating procedures. The value of U was computed by the simple expedient of subtracting every score in the first sample from every score in the second sample and counting the number of minus signs which resulted. This number is U by definition. The possibility for a tie was ignored since the expected rate of ties was approximately one per thousand Us, a rate which, while significant, would seem to have little effect on the observed results. As in the t program, the obtained U values were sorted and tallied on an internally contained table which was punched out in card from when the desired number of Us from the specified populations was reached.

RESULTS

Since the results of the study will be expressed in terms of empirical power functions, the investment of a small amount of space to elaborate on the method of their determination may be in order at this point.

For any given set of conditions, i.e., combination of means, variances, and distributions, the result of the computer procedure is a set of t's or Us which may be arranged in a frequency distribution. Figure 1 shows two such distributions. The distributions are of t's obtained on the basis of sample sizes of 15 from normal populations having a variance of 1.0. One of the distributions shown, that centered around the t value of zero, arose when the means of the two samples were both equal to zero. For the other distribution, that centered around 2.8 and shaded, the difference between the means of the two samples was equal to 1.0, that is, equal to one standard deviation. The vertical lines divide the range of possible t values into two regions. The unhatched region marks off those values of t which result in a decision to *accept* (or fail to reject) the null hypothesis at the .05 level. The hatched regions, those values of t which result in a decision to *reject* the null hypothesis at that level. As can be seen, when the null hypothesis is true most of the obtained t's fall in the region of acceptance. The proportion which fall in the region of rejection, in this case .049, we have called the empirical alpha level.

Sometimes a decision is made to accept a false null hypothesis, the so-called beta error. In Figure 1, such an error would be made in the case of those values of t from the shaded distribution which fall in the region of acceptance. On the other hand, decisions to reject the null hypothesis when it is indeed false occur for that proportion of the t's from the shaded distribution which fall in the region of rejection, in this case .737. It is this pro-

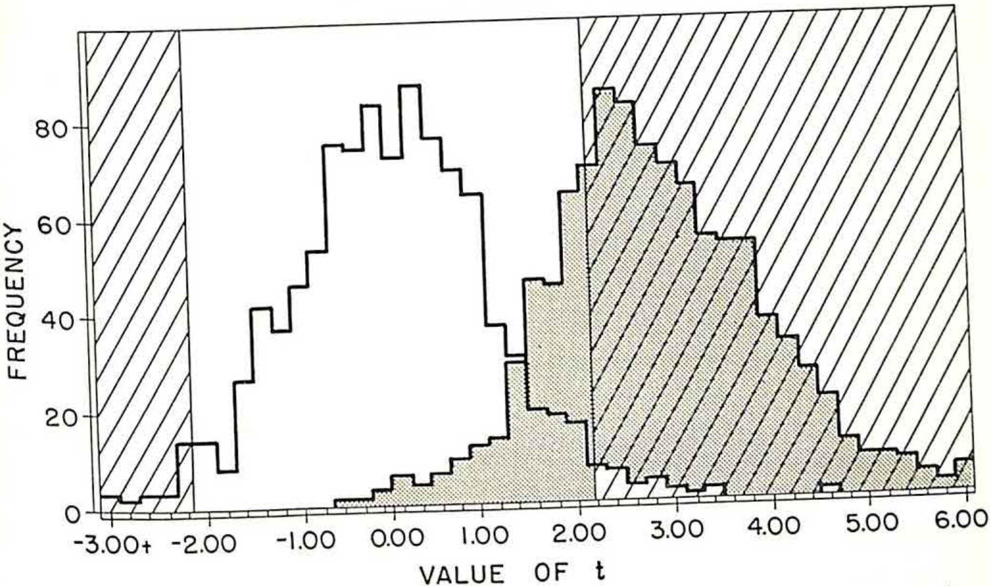


FIG. 1. Empirical *t* distribution for mean difference of (*a*) 0.00 (unshaded area) and (*b*) 1.00 (shaded area) for sample sizes of 15. (The hatched area is the region of rejection for the .05 level of significance.)

portion which we call the empirical power of the test under the given conditions. It should be noted that both the empirical alpha and the empirical power are estimates of the exact theoretical values which obtain for these conditions.

The data of the present study are empirical power values considered as functions of the actual difference between population means. We shall consider separately, the functions for the one- and two-tailed tests and for the .05 and .01 values of alpha. The figures which will be presented will, for a given set of conditions, depict such functions for both that *t* and *U* tests so that visual comparisons may be made.

One further thing should be said about nomenclature before results are presented. As in the previous paper, the conditions of sampling will be symbolically represented. For example, N(2,1)5-N(0,1)15 indicates that the first sample is from a normal population N with a mean of 2 and a variance of 1, the sample size being 5. In this instance, the second sample is from a normal distribution having a mean of 0, a variance of 1, and the sample size is 15. In the text and figures, we will use the variable "x" for the value of the first mean to indicate that it takes on the several values necessary for the $\mu_1 - \mu_2$ values on the abscissa. In all cases the value of the mean of the second distribution is zero. One thousand *U*s and *t*'s were obtained for each condition.

Normal Distribution: Homogeneous Variances

First to be considered are the cases in which the sizes of both samples are the same. Figure 2 depicts the various empirical power functions obtained when the condition of sampling is N(x,1)5-N(0,1)5. Because of the discrete nature of the *U* distribution, exact alpha values of .05 and .01 are not possible. When both samples sizes are 5, the following alpha values obtained from tables are used: .056 for the two-tailed test and .048 for the one-tailed test in place of .05, and .008 for both the one- and two-tailed test in place of .01.

A number of interesting facts can be observed in Figure 2. First is the expected superiority of the one-tailed test to the two-tailed test. Of prime importance for the present paper is the remarkable lack of difference in power of the *t* and *U* tests over the range of values

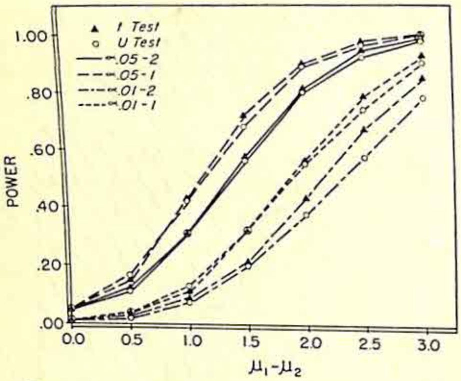


FIG. 2. Empirical power functions for U and t tests with sampling scheme N(x,1)5-N(0,1)5.

presented. Only in the case of the two-tailed at the .008 level does the t test seem to exhibit a definite superiority and here only for gross difference between the means. In most of the other cases, the differences between the obtained points are well within the margin of sampling error (i.e., not significantly different). This would indicate that although the t test is theoretically a uniformly (over all mean differences) most powerful test, the margin over the U test is not very much. Because of this power property of the t test under these conditions, the points at which the U test shows a superiority must be attributed to sampling error. It is also clear that the alpha values ($\mu_1 - \mu_2 = 0$) are virtually identical and are approximately the theoretical values expected.

Increasing both sample sizes from 5 to 15 tends, as shown in Figure 3, to increase the power of the test, but to leave virtually unaffected the things which we have said about the smaller sizes. For example, at the .05 level for a one-tailed test, the power of both tests for a difference between means of 1.0 is around .42 for sample sizes of 5, and .85 for samples of size 15.

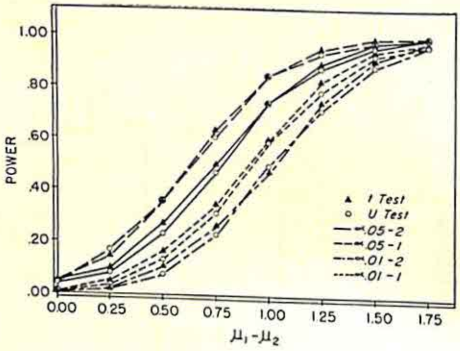


FIG. 3. Empirical power functions for U and t tests for sampling scheme N(x,1)15-N(0,1)15.

Still maintaining the condition of sampling from a normal distribution with equal variances of the parent populations, we may allow the sample sizes to be different, 5 as opposed to 15, and generate another series of curves, Figure 4. The results are similar to those preceding, with the power for the one-tailed test at the .05 level and a mean difference of 1.0 being in this case about .60. This, too, is to be expected since the power of the test among other things is an increasing function of the difference between the means but a decreasing function of the expected standard error of the difference between means where

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$$

Figure 4 shows more clearly than any of the preceding graphs the superiority of the t test to the U test, but the largest obtained difference is only .12, and being the largest probably overestimates the true difference.

The results we have observed thus far are for conditions in which the assump-

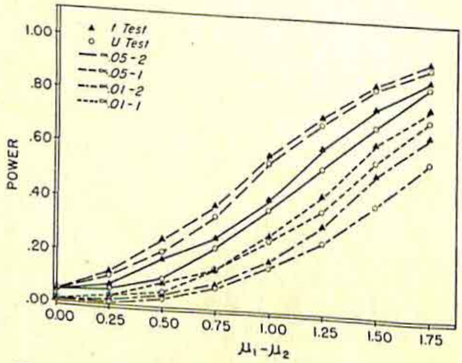


FIG. 4. Empirical power functions for U and t tests for sampling scheme N(x,1)5-N(0,1)15.

tions underlying the *t* test are satisfied. Under these conditions we know that theoretically the *t* test is the most powerful test of the difference between two means. The power functions and relations between the powers of the two tests obtained in this study are comparable with those found in several sources (Dixon, 1954; Dixon & Massey, 1957; Ferris, Grubbs, & Weaver, 1946, to name a few). Thus they contribute only a confirmation of the general method and at the same time furnish a graphical demonstration of the fact that, while less powerful than the *t* test, the *U* test performs quite well in those situations for which the *t* test is expressly suited.

When the assumptions of the *t* test are not fulfilled, it is not necessarily a most powerful test. Moreover, as stated earlier, there are theoretical conditions for which the *t* test is considerably less powerful than other tests including the *U* test. The remainder of this paper will compare power functions for the *U* and *t* tests for those violations of assumptions which arise from various combinations of the three distributions, the two variances, and the two sample sizes which have been selected for the study. We will determine whether the power functions for the *t* test show any drastic deviations from those power functions we have already seen, and we shall discover those cases, if any, for which the *U* test performs better than does the *t* test. This, of course, must be within the limitations imposed by the selection of conditions.

Normal Distributions: Heterogeneous Variances

Initially, we shall proceed by examining the effect of violating the assumption of homogeneity of variances with normal distribution. In the previous study, it was determined that inequality of variances up to at least a ratio of 1 to 4 produced a very little effect in alpha provided the sizes of the two samples are the same. If the sample sizes are different, gross disturbances in alpha occur.

Figure 5, depicting the equal-sample-size case ($n = 5$), reveals that for *t* the alpha is relatively undisturbed and power also seems little affected. It should be noted that the relatively low power for Figure 5 (and for some following figures) for a given mean difference is to be attributed to a change in the standard error of the difference between means occasioned by the change in σ_2 from 1 to 4. For a given mean difference in standard error units, the power remains

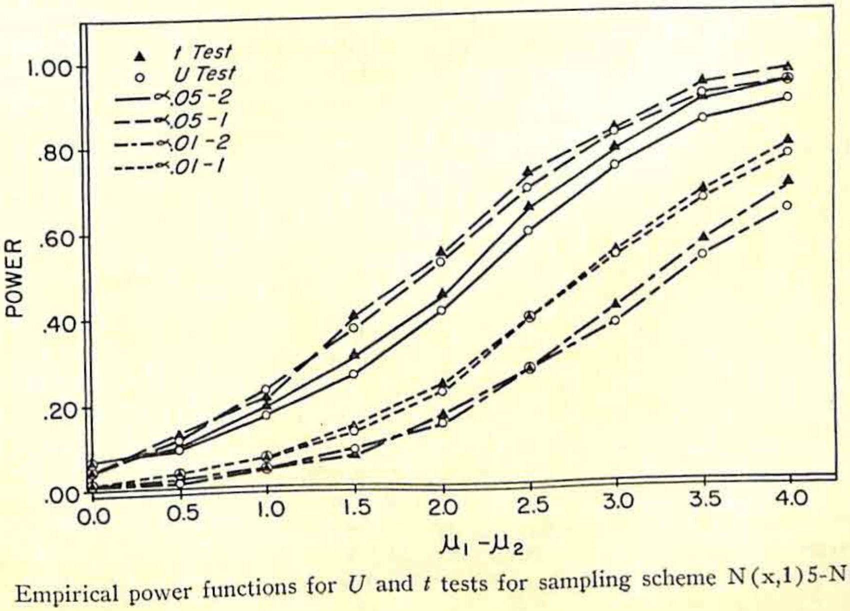


FIG. 5. Empirical power functions for *U* and *t* tests for sampling scheme N(x,1)5-N(0,4)5.

essentially the same. For example, the true standard error of the difference between means for the case N(x,1)5-N(0,4)5 is .63, while for the case (x,1)5-N(0,4)5 it is 1.0. A mean difference of 2.0 (2 standard errors) in Figure 5 shows the power of the two-tailed *t* test at the .056 level to be approximately .45. On Figure 2, a mean difference of 1.26 (2 standard errors) shows the comparable power to be approximately .44. In terms of the true standard error of the difference between sample means, then, the power of the *t* test seems relatively unaffected by violating the homogeneity of variances assumption.

Likewise, the power of the *U* test is maintained, but at a level again slightly less powerful than the *t* test. Similar results (not shown) obtain when both sample sizes are changed to 15.

Introducing heterogeneity of sample size as well as heterogeneity of variance produces discrepancies which might be predicted from what is already known about the effect of this condition on the alpha level. Figures 6 and 7 portray these effects for the two possible combinations of variance and sampe size considered, N(x,1)5-N(0,4)15 and N(x,4)5-N(0,1)15 respectively. The first thing to observe in these figures is that effect on the alpha level (the power for a mean difference of zero). In Figure 6 the alpha is for all curves less than the nominal values of .05 and .01, although the magnitude of the difference for the *U* curves is less than those for *t*. The actual values observed are as follows: (one-tailed test-.05 level) $t = .010$; $U = .029$; (two-tailed test-.05 level) $t = .009$; $U = .027$; (one-tailed test-.01 level) $t = .002$; $U = .005$; (two-tailed test-.01 level) $t = .001$; $U = .004$. The power functions reflect the fact that while nominally at a .05 level, the tests are actually operating at a reduced alpha value. At the reference value, a mean difference equal to two standard errors, the power of the two-tailed .05-level *t* test is approximately .24, while the power for the comparable *U* test is about .28. Both of these values are much less than the values of about .45 which obtained under the other conditions. We may conclude that the conditions of heterogeneity of variances and sample sizes which produced Figure 6 have affected the alpha level of both *U* and *t*, *U* to a lesser extent, however, than *t*. Since the alpha level for the *U* test under these conditions is greater than that for the *t* test, the resulting power functions should and do reflect a superiority of *U* to *t*. What we are observing is the power curve for *t* for an alpha level of about .01 and for *U* of about .03 when we consider, for example, the curves which were constructed on the basis of the normal boundaries of the region of rejection for the .05 level of alpha.

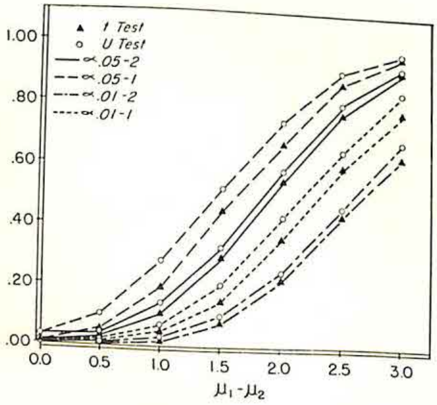


FIG. 6. Empirical power functions for *U* and *t* tests for sampling scheme N(x,1)5-N(0,4)15.

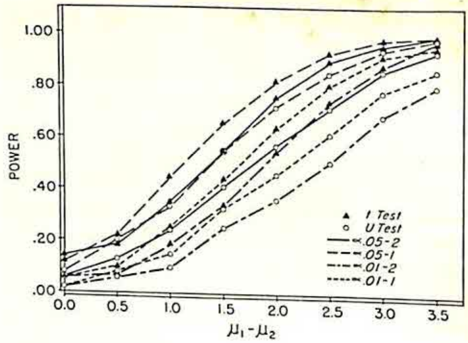


FIG. 7. Empirical power functions for *U* and *t* tests for sampling scheme N(x,4)5-N(0,1)15.

We may make exactly the same comparisons for the results depicted in Figure 7, but note that the effects are in the opposite direction to those found in Figure 6. For example, the alpha values obtained are all greater than the nominal values, and again the distortion to the t values is greater than those for U. For this case—N(x,4)5-N(0,1)15—the actual alpha values observed are: (one-tailed test-.05 level) $t = .115$; $U = .081$; (two-tailed test-.05 level) $t = .145$; $U = .064$; (one-tailed test-.01 level) $t = .048$; $U = .020$; (two-tailed test-.01 level) $t = .058$; $U = .021$. Considering the reference point used earlier, the empirical power of the two-tailed test with alpha equal to .05 for a difference between means equal to two standard errors of the difference between means, we find these values to be .75 for t and .52 for U. And as before, the power curves behave as if they were the power curves for the observed value of alpha, the curves for t being considerably above those for U at all points.

We may conclude that violating the assumption of homogeneity of variance if the underlying distributions are normal has little effect on either the alpha level or the power of the t or the U test as long as sample sizes are the same. The violation of this assumption coupled with heterogeneous sample sizes changes the alpha level of both the t and the U tests and produces power functions which seemingly are roughly appropriate for the true alpha level rather than the nominal one. The U test seems much less disturbed by this particular violation than does the t test, but it is by no means true that the U test is completely unaffected as would seem to be implied by the term "nonparametric." Rather it seems to behave much as does the t test, but somewhat less sensitively to the violation of the assumption of homogeneity.

Nonnormal Distributions

At this point we will examine the empirical power functions for the t and the U tests when sampling for at least one sample is from nonnormal distributions. In this way we will observe the effects, if any, on the functions if one or both of the parent populations is exponential, or if one or both of the parent populations is rectangular. From an examination of the empirical alpha for the t test, we already know what the effects of sampling from a rectangular distribution are minimal on alpha. When the exponential distribution is involved, however, some perturbations in alpha may occur because of differences in skewness of the two distributions. The magnitude of these disturbances as seen in the earlier study was progressively lessened as sample size increased. At sample sizes of 25, the effect was virtually unnoticeable.

These earlier observations are confirmed in this study. Figure 8, for example, depicts the results of sampling from two rectangular distributions—R(x,1)5-R(0,1)5. Alpha is approximately the correct value for both levels for both tests. In this figure, it appears that the power of t is quite generally greater than that for U, but never by much except for the .01 level. Indeed, it would seem that for small differences between means, say, 0.5, U may be superior to t. Similar results (not shown) were obtained when both sample sizes were increased to 15.

The next distribution to be examined occurs when both samples are taken from exponential distributions as in Figure 9—E(x,1)5-E(0,1)5. The empirical alpha values are in the appropriate

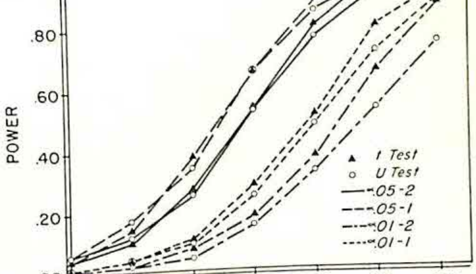


FIG. 8. Empirical power functions for U and t tests for sampling scheme R(x,1)5-R(0,1)5.

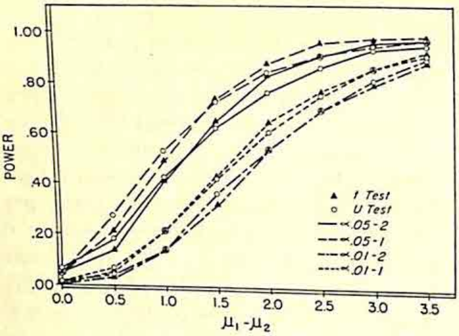


FIG. 9. Empirical power functions for U and t tests for sampling scheme E(x,1)5-E(0,1)5.

ranges. Here again, for small differences between the means, the U test seems consistently more powerful than the t test, but this advantage disappears with greater mean differences. All in all one may conjecture that when both distributions have the same shape, even though not normal, the power functions of the t and U tests have a relatively constant relationship. In most instances, the t test is slightly more powerful than the U test, as we have observed in most of our examples.

All of the foregoing power functions have come from combinations of distributions which produce essentially symmetric distribution of t and U when the null hypothesis is true (zero difference between means). We might expect more severe disturbances in power functions in those cases which, because of basic asymmetries in the observed t distributions, exhibited discordant obtained alpha values in the previous study. As may be recalled, the asymmetric distributions arose when the two parent populations differed in skew, one normal, the other exponential, for example. It was true, however, that increasing the sample size greatly diminished the asymmetries since the underlying probability mechanisms tend to normalize the distribution of t as sample size increases.

It is such conditions, probably not too uncommon in the experience (or at least the imagination) of the research worker, which motivate a desire to seek statistical tools which exhibit fewer allergic reactions to violations of assumptions.

Because of these considerations, it is interesting to compare the power functions of the t and the U test under such conditions of sampling. Figure 10 makes the comparisons for the case of exponential and normal distributions with samples of size 5—E(x,1)5-N(0,1)5. First to be noted are the discrepancies in the obtained alpha values for the one-tailed tests due to the asymmetrical distribution of t and, surprisingly enough, for U. The .048-one-tailed values are .022 for U, .014 for t. Likewise, the .008-one-tailed values are .002 for U, .001 for t. This is to be contrasted with the two-tailed values which are relatively close to the nominal values—the .056-two-tailed value for U is .049, for t it is .068; and the .008-two-tailed value for U is .014, for t, .011. These results for t are comparable to those found in the previous study. Thus we find that when considering alpha, t and U both are affected by sampling from populations which differ in skew, although it is possible that the effect on U is less than that on t.

A further examination of Figure 10 reveals that the empirical power functions for U are, with one or two exceptions, higher than the power functions for t, but the advantage is slight.

As mentioned previously, increasing the sample size tends to normalize the

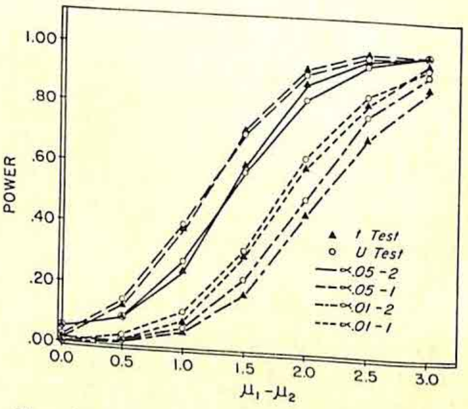


FIG. 10. Empirical power functions for U and t tests for sampling scheme E(x,1)5-N(0,1)5.

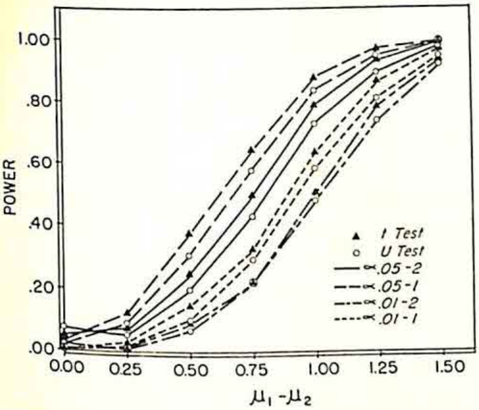


FIG. 11. Empirical power functions for *U* and *t* tests for sampling scheme E(x,1)15-N(0,1)15.

distribution of *t* in the null case, the effect being to lessen the discrepancy of nominal and obtained values for the one-tailed test. Figure 11 presents the curves for the increased sample size case—E(x,1)15-N(0,1)15. As before, our attention is first directed to the zero-mean-difference points, and as expected, the values for *t* have become closer to the nominal values—one-tailed test, .037 for .05 level, and .006 for .01 level; two-tailed test, .044 for .05 level and .012 for .01 level. Increasing the sample size to 15, however, does not seem to improve the performance of the *U* test. For the one-tailed *U* test, the obtained alpha values are .016 for the .05 level, and .000 for the .01 level. For the two-tailed test, the obtained alpha values again seem to be higher than the nominal values—.068 for the .05 level, and .019 for the .01 level. In Figure 11, we note that the power functions for the *U* test are almost invariably below those for the *t* test.

A notable phenomenon occurs for the .05-level, two-tailed *U* test. This particular curve shows a decrease in power from 0 mean difference to 0.25 before starting up again. While sampling error may well account for the dip, it is certainly possible for a test to be "biased," the technical term for such an occurrence. In fact there is no reason to believe that the *U* test as a test for mean differences should not be biased. The *U* test is only fortuitously a test for mean differences, being fundamentally a test for differences between distributions. As we have seen it is less sensitive to some kinds of differences between distributions than others, being perhaps maximally sensitive to differences in central tendency of two distributions. But there are many measures of central tendency, the mean and the median, for example. We know that the mean and median are different in skewed distributions. In the combination of distributions with which Figure 11 is concerned, namely, exponential and normal, it is possible for medians to be the same for the two distributions which necessitates that the means be different and vice versa. For the *t* test which explicity evaluates differences in sample means this seems to present no problem. The *U* test, however, may well be more sensitive to median differences and thus show the bias when used as a test for differences between means.

Other combinations of distributions which were explored in no way change the general picture which we have continually observed. We may summarize that picture as follows: In general the *t* test is more powerful than the *U* test, but never by much. Based on the evidence we have seen, one might conjecture that over a long series of experiments involving distributions of the kind we have used, a consistent use of the *t* test might result in, say, 5% more rejections of a false null hypothesis than would a consistent application of the *U* test. There are many other kinds of distributions arising in research, however, for which this statement need not apply. Depending upon one's inclinations, such a conclusion could be interpreted as ample grounds for habitual use of either of the two tests. There are, as we have seen, other considerations. As a general rule we might say that the *t* test seems to provide the appropriate power curve for the actual alpha level involved, whereas, the *U* test shows more variability in its power functions. That is to say that if the actual (not the nominal) alpha is .05, the power curve for the *t* test in most cases is probably very similar to the power function for the theoretical case for the .05 level; it is also to say that the statement is not nearly so true of the U test. This property of the t test is useless, however, unless we know what the actual alpha is. In cases when assumptions are violated we will not. It is true that the violations of the assumptions underlying the t test can produce large discrepancies between nominal and actual alpha when the sample sizes are other than nearly equal. This in itself contraindicates the use of the t test in these instances. A further somewhat surprising consideration is that the U test is not truly distribution free. It is always sensitive to differences in distributions, and sometimes seems more affected by differences (other than mean differences) than is the t test.

If one final word is to be said it might be that one should not avoid using the t test (provided relevant considerations have been made) solely on the grounds that it is subject to error when assumptions are violated and that the U test is not subject to error under the same conditions. Both of the statements are unreasonable in view of the data. One should not, however, refrain from using the U test in place of the t test on the grounds that it is considerably less powerful than the t test. This is simply not true.

REFERENCES

AUBLE, D. Extended tables for the Mann-Whitney statistic. *Bull. Inst. Educ. Res., Ind. U.*, 1953, **1** (2).

BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, **57**, 49–64.

DEUCHLER, G. Über die Methoden der Korrelationsrechnung in der Pädagogik und Psychologie. *Z. Pädag. Psychol.*, 1914, **15**, 114–131, 145–159, and 229–242.

DIXON, W. J. Power under normality of several nonparametric tests. *Ann. math. Statist.*, 1954, **25**, 610–614.

DIXON, W. J., & MASSEY, F. J. *Introduction to statistical analysis.* (2nd ed.) New York: McGraw-Hill, 1957.

FERRIS, C. D., GRUBBS, F. E., & WEAVER, C. L. Operating characteristics for the common statistical tests of significance. *Ann. math. Statist.*, 1946, **17**, 178–197.

HODGES, J. L., & LEHMAN, E. L. The efficiency of some nonparametric competitors of the *t*-test. *Ann. math. Statist.*, 1956, **27**, 324–335.

KRUSKAL, W. H. Historical notes on the Wilcoxon unpaired two-sample test. *J. Amer. Statist. Ass.*, 1957, **52**, 356–360.

LEHMAN, E. L. The power of rank tests. *Ann. math. Statist.*, 1953, **24**, 23–43.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

MOOD, A. M. On the asymptotic efficiency of certain nonparametric two-sample tests. *Ann. math. Statist.*, 1954, **25**, 514–522.

SIEGEL, S. *Nonparametric statistics: For the behavioral sciences.* New York: McGraw-Hill, 1956.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics*, 1945, **1**, 80–83.

(Received June 29, 1961)

VOL. 69, NO. 4 JULY 1962

PSYCHOLOGICAL REVIEW

CHANGES IN THE APPEARANCE OF STIMULI OF VERY HIGH LUMINANCE[1]

TOM N. CORNSWEET
University of California, Berkeley

When quanta of light strike the retina, some of them are absorbed by the pigment molecules contained in the visual receptor cells, and this absorption ultimately results in the activation of the visual nervous system. Recent developments in the biochemistry of the visual pigments has given us a fairly clear picture of the microstructure of quantal absorption. A theoretical analysis of this first stage in the visual process will be presented and this analysis will be shown to furnish a possible explanation of certain striking changes in hue and brightness that occur when a stimulus of very high luminance is steadily viewed.

THEORY OF VISUAL PHOTOCHEMISTRY

There is fairly good agreement that light triggers neural events by processes of the following general type. Quanta of light, after passing through the other media of the eye, arrive at the retina where they may or may not hit receptor cells. Those quanta that finally arrive in the region of the visual pigment of the receptor cells may or may not be captured by the molecules of pigment. If a quantum is in the close vicinity of a particular pigment molecule, the probability that it will actually be captured by the molecule depends upon the relationship between the wave length, or energy, of the quantum and the physical structure of the molecule and its surroundings. If all these probabilities are satisfied, the quantum is absorbed by a visual pigment molecule. When, before absorption of the quantum, the pigment molecule was in a state that will hereafter be called the "regenerated" state, the absorption will cause the molecule to change its state (to what will be called an "unregenerated" state). For example, Kropf and Hubbard (1958) suggest that the change in state is a cis-trans isomerization. The changing of state of a regenerated pigment molecule sets up a process in the receptor which eventually results in the discharge of a packet of some chemical mediator at a synapse farther along in the retina. When the concentration of chemical mediator at the synapse is great enough, a nerve impulse is triggered, and either that impulse or another resulting from it, travels up the optic nerve. At the synapse, there is also a substance that destroys the chemical mediator and this, combined with diffusion effects, give the mediator a short "half life." Therefore the probability of triggering a nerve impulse depends upon the *rate* of discharge of packets. Above threshold, the frequency of nerve impulses will be greater, the greater is the rate at which packets of mediator are being discharged. Since each discharged packet is the ultimate result of the absorption of a quantum by a regenerated molecule, the frequency of nerve impulses will depend upon the rate at which quanta are absorbed by regenerated molecules.

---

[1] This research was supported by United States Public Health Service Grant Number B2169. The writer also wishes to thank Robert Abelson and Norman Anderson for their helpful suggestions, and James Thomas for his work on the brightness reversal experiment.

The changes that occur in retinal photopigment when light is absorbed are reversible, and, if the eye remains in darkness long enough, virtually all of the pigment will have regenerated to its initial, ready-to-trigger, state. Some of the more recent evidence on cone pigments *in vivo* indicates that at least the main part of this regeneration process goes on at the same rate whether the eye is in darkness or in light, the rate depending only upon the concentration of unregenerated pigment present (Rushton, 1958).

If the above general conceptions are correct, then, when a subject is steadily fixating a field of fixed luminance, retinal pigment is both being broken down and regenerated, and the frequency at which nerve impulses are being fired in any one optic nerve fiber depends upon the rate at which regenerated molecules are being broken down. Hereafter, the term "output" will be used to denote the instantaneous concentration of chemical mediator at the receptor terminus. This receptor output is thus directly related to the rate at which quanta are absorbed by regenerated molecules, and it is similarly related to the frequency of triggering of nerve impulses somewhere farther along the retinal activity chain.

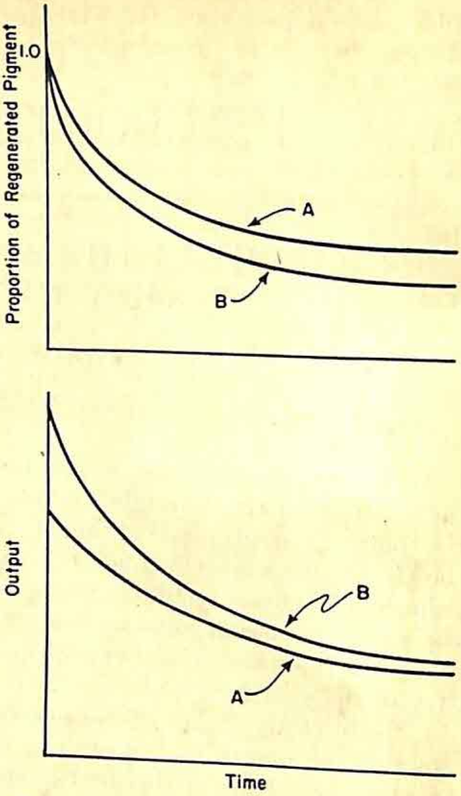


FIG. 1. Proportion of regenerated pigment as a function of the time after onset of the stimulus, for two different illuminances (B greater than A); and output as a function of the time after onset of the stimulus for the same two illuminances.

The upper curve, labeled A, in the upper part of Figure 1 is a generalized representation of the changes in pigment concentration that would occur during stimulation of a single receptor if the preceding considerations are correct. The receptor is completely dark adapted before the onset of the stimulus, so that the proportion of regenerated to total pigment is 1.0. At the time represented at the origin, a steady light is turned on. The pigment concentration will drop rapidly at first and then level off as it approaches some asymptotic level, the level at which the rates of breakdown and regeneration are equal. The lower curve, labeled B, represents the same events for an identical receptor under a more intense stimulus.

The curves in the lower part of Figure 1 are a generalized representation of the outputs (rates of pigment breakdown) of the same two receptors under the conditions of the upper part of Figure 1. Since the receptors are taken to be identical, and since at time 0 both concentrations are 1.0, the initial output of B will be greater than the initial output of A. Further, since the time constants of regeneration for A and B are taken to be equal, the output of B must approach a final level that is higher than the asymptote of A. This is true even though the equilibrium concentration of regenerated pigment is lower for B than for A.

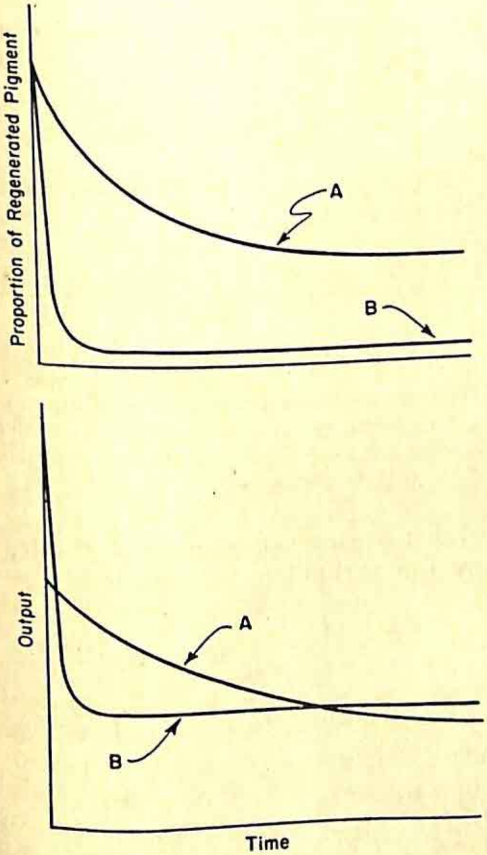


FIG. 2. Proportion of regenerated pigment as a function of time after onset of the stimulus, for two different illuminances (A at moderate illuminance and B approaching an infinite illuminance); and output as a function of time after onset of the stimulus for the same two illuminances.

The upper part of Figure 2 represents the regenerated pigment concentration curves for the same two receptors when the illuminance on B has been raised almost to infinity, but the illuminance on A is the same as in the preceding figure. In the first short interval after the light is turned on, virtually all of the pigment molecules in B will be broken down. From that time on, each molecule that flips back to the regenerated state will immediately be broken down again. In other words, when the illuminance approaches infinity, the concentration approaches equilibrium extremely rapidly, and the equilibrium level depends only upon the time constant of regeneration of the pigment. The lower part of Figure 2 shows the output curves for these receptors, again when the illuminance on B approaches infinity. It will again be true that, when the stimuli are first turned on, B will have the higher output. Further, at equilibrium, B must still have the higher output, even though its regenerated pigment concentration is very much lower. In general, the equilibrium rate of breakdown of regenerated pigment (output) is greatest when the illuminance is infinity, since that rate equals the rate of regeneration and since the rate of regeneration will be greatest when the concentration of *unregenerated* pigment is greatest. But, since

Receptor B will reach its equilibrium level almost instantaneously, its output curve must cross the output of A twice, as in the lower part of Figure 2. (This is necessarily true as long as the illuminance on A is great enough that its initial output level is greater than the equilibrium level for B.) Obviously the illuminance on one of a pair of receptors need not be nearly infinite in order that the pair exhibit a reversal of output relations. There will be a large set of pairs of illuminances for which reversal should occur.

A MORE SPECIFIC SET OF ASSUMPTIONS AND THEIR CONSEQUENCES

Now let us substitute, for each of the general conditions described above, a corresponding specific assumption about the operation of the visual system. From these assumptions, certain quantitative aspects of the visual phenomena to be discussed may be derived, and the derivations examined in comparison with the data themselves.

If the absorption of a quantum by a particular molecule is independent of the state of the other molecules in the receptor, then it follows from statistical considerations that the rate of absorption by regenerated molecules is:

$$\left(\frac{dN}{dt}\right)_- = AQN \qquad [1]$$

where N is the number of regenerated molecules, A is the probability that a quantum incident on a region containing a regenerated molecule will bleach that molecule (the photosensitivity of each regenerated molecule at the wavelength of the incident quantum), and Q is the number of quanta incident during the Time dt.

Recent studies of the bleaching of cone pigments in the living human eye support the notion that, at least to a good approximation, the regeneration of cone pigments is a first order chemical reaction (Rushton, 1958). We will assume, accordingly, that the probability that any particular unregenerated molecule will be regenerated in a unit of time will be independent of the states of the surrounding pigment molecules. It follows then that the rate of regeneration of unregenerated molecules is:

$$\left(\frac{dN}{dt}\right)_+ = K(M - N) \qquad [2]$$

where M is the total number of molecules, regenerated or not, in the receptor, N is, again, the number of regenerated molecules, and K is the probability that an unregenerated molecule will be regenerated in Time dt.

The net rate of change of regenerated molecules is:

$$\frac{dN}{dt} = K(M - N) - AQN \qquad [3]$$

We will now define the quantity X, the proportion of regenerated pigment in a receptor:

$$X = N/M \qquad [4]$$

Substituting Equation 4 into Equation 3 and dividing through by M:

$$\frac{dX}{dt} = K(1 - X) - AQX \qquad [5]$$

The solution to this differential equation yields:

$$X = \frac{K}{K+AQ} + \frac{AQ}{K+AQ} \cdot e^{-t(K+AQ)} \qquad [6]$$

when the initial condition is $X = 1$, that is, complete dark adaptation.

In the preceding section, the output of a given receptor was defined as the rate at which quanta are absorbed by regenerated molecules. Therefore:

$$\Omega \equiv \left(\frac{dN}{dt}\right)_{-} = AQN = AQMX$$

$$\Omega = AQM\left[\frac{K}{K+AQ} + \frac{AQ}{K+AQ}\cdot e^{-t(K+AQ)}\right] \quad [7]$$

This is an equation that relates the output of any given receptor to the irradiance and duration of a steady light falling upon it.[2] The lower parts of Figures 1 and 2 are actually plots of Equation 7 for various values of Q.

CONSEQUENCES FOR BRIGHTNESS

It seems to be a reasonable assumption that the brightness of a lighted area will increase as the frequency of nerve impulses originating from the retinal region corresponding to that area increases. If this assumption and those concerning the conversion of quanta to impulses are correct, then the falling of the curves in the lower parts of Figures 1 and 2 represent the fact that the brightness of a stimulus of fixed luminance decreases as the eye light adapts to it.

The considerations illustrated in the lower part of Figure 2 also lead to the following prediction. If the eye is first dark adapted, then one area of the retina is illuminated (say with white light), and another area is simultaneously illuminated much more intensely, the more intensely illuminated area will look first brighter than the other, then darker, and finally somewhat brighter again. That is, there should be a period of time during which the less intensely illuminated region should look brighter than the more intensely illuminated one. This paradoxical prediction has been tested and it is correct. The predicted brightness reversal does indeed occur, and is very easy to demonstrate.

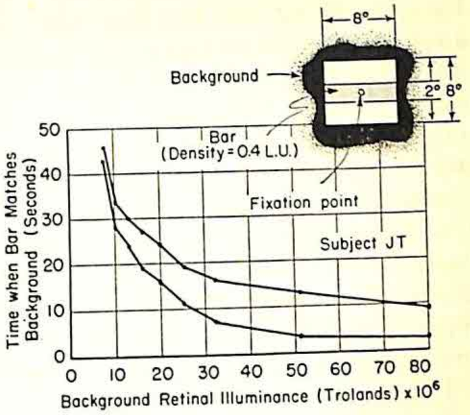


FIG. 3. Times of brightness reversals as a function of retinal illuminance for the stimulus configuration shown in the upper right corner. (The surround was dark, the background illuminated, and the horizontal bar consisted of a strip of 0.4 log unit neutral density filter. Each point is a mean of four judgments.)

The data in Figure 3 were taken under the following conditions. The subject fixated a small bright point in the center of the field shown in the upper right corner of the figure. The field was shown in Maxwellian view, through a simple optical system whose effective aperture was 3.5. The source was a 32-candle power automobile taillight bulb, run from a regulated DC supply. The filament image at the pupil was 1 millimeter in diameter, centered on the natural pupil. The luminance of the field was varied with neutral density filters. Heat glass was used to filter out the infrared.

---

[2] None of the steps in this derivation is original. Pieces of it can be found in the writings of Troland (1930), Hecht (1935), Rushton (1958), for example. And, although the writer has not found the complete derivation all in one place in the literature, it may well be there somewhere.

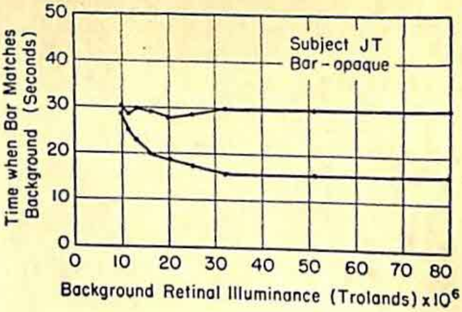


FIG. 4. Times of brightness reversals as a function of retinal illuminance. (For these data, the bar, as shown in Figure 3, was replaced by an opaque strip. Each point is the mean of four judgments.)

When the field is first turned on, it looks like a dark bar on a bright field. But after a time that depends upon the luminance relations, the field abruptly reverses, so that the bar looks considerably brighter than the background. Then the brightnesses gradually converge and reverse again, until finally the background looks just slightly brighter than the bar.

To collect the data in Figure 3, the following procedure was used on each trial. The subject's viewing eye was dark adapted for 5 minutes. Then the stimulus figure was turned on and two clocks started. The subject looked steadily at the fixation point, and at the first instant when the bar and its background were equally bright (changing from normal to reversed brightnesses) he pressed a key, stopping one of the clocks. When the brightnesses were again equal (changing from reversed to normal) he pressed a second key, stopping the second clock. The two times were recorded, and each point on the curves in Figure 3 is a mean of four such times. The area between the two curves in Figure 3 represents the times when the brightnesses are reversed. At least 45 minutes elapsed between successive trials, and the order of illuminances was balanced.

The second equality point is much harder to judge than the first one, and the judgments are less reliable. Part of the reduction in reliability is due to the fact that the changes are occurring very gradually, but, in addition, strong hue differences between the bar and ground appear after the first 20 seconds or so; the background turns reddish and the bar less red. (There is not much point in trying to interpret these color differences physiologically, because the wave length composition of the stimuli in this set of experiments was very complex, due to absorption by the heat glass, etc.)

The brightness reversal occurs equally strongly whether the fixation point is in the bar or the background. It cannot be explained by inhomogeneities in the retinal structure.

The data in Figure 3 were taken when the bar was .4 log units less intense than the background. (The bar was simply a strip of .4 density neutral gelatin filter pasted across the larger square aperture that formed the background.) But brightness reversal also occurs when the bar transmits no light to the retina. The data for Figure 4 were taken after the strip of .4 density filter was replaced by an opaque strip. In this case, the time relations are less strongly dependent upon the illuminances (and the hue differences more extreme), but the effect is still very strong. The fact that the strip is opaque does not mean that the actual retinal illuminance in the region of the image of the bar is zero. There is extensive evidence in the literature to indicate that the stray light in this sort of stimulus situation is appreciable (see DeMott & Boynton, 1958, for an extreme example). Furthermore, there is a good possibility that the illuminances at

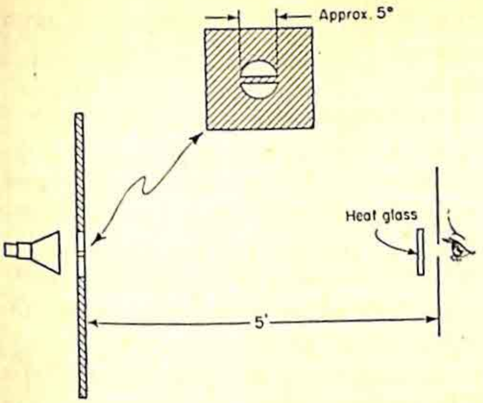


FIG. 5. Apparatus for a demonstration of brightness reversal. (The inset figure shows the subject's view.)

the edges of the figure (where stray light is greatest) are the critical ones. (This possibility is currently being tested.) Thus, the retinal condition in which an opaque bar is used in the apparatus is on a continuum with any in which a filter of transmittance greater than zero is used.

It is easy to set up a demonstration of this brightness reversal, as illustrated in Figure 5. A strip of cardboard is placed across a round hole in a larger board. This hole is just a little smaller than the face of the sort of photoflood lamp that has a frosted face and a built-in reflector (e.g., GE #PH/RFL2). A small hole in the middle of the strip will serve as a fixation point (it is essential to fixate very steadily for this demonstration). The face of the photoflood lamp should be viewed through the hole and strip, and through heat absorbing glass, from about 4 or 5 feet away.

AN EXTENSION OF THE THEORY TO CHANGES IN HUE

When a subject steadily fixates a field of extremely high luminance in the long wave length region of the spectrum, the hue of the field undergoes very striking changes. Some of these changes have been mentioned independently by Auerbach and Wald (1955) and by Noddack and Jarzyk (1955), in the course of studies of other color phenomena. Cornsweet, Fowler, Rabedeau, Whalen, and Williams (1958) briefly reported a study of the color changes themselves. According to all of those articles, the nature of the color changes is as follows. The subject looks steadily at a field that is illuminated with light from the long wave length end of the spectrum. When the luminance is low or moderate, the field will look red or orange. However, if the luminance is extremely high, it will look orange at first, but then rapidly shift in hue through yellow to a vivid green. This pattern of changes is reported to occur for stimuli of several different wave lengths in the "red" end of the spectrum, and for monochromatic as well as broad band stimuli. Stimuli with wave lengths in the green region of the spectrum are reported to produce extreme desaturation, but no clear hue changes are reported. Auerbach and Wald (1955) also report that bright blues will change to red, but such a change was not observed in either of the other two studies cited. A review article by Cohen (1946) provides a number of references to other studies of the effects of high luminance stimulation.

The theoretical considerations discussed in the preceding sections suggest an explanation for some of these changes in hue. The explanation is based upon the retinal mechanisms already mentioned, along with some simple and widely held assumptions about the physiological bases for color vision.

First, we will assume that there are at least two sorts of cones that differ in the pigments they contain, the two pigments having different but overlapping absorption spectra. The absorption spectra for such a pair of

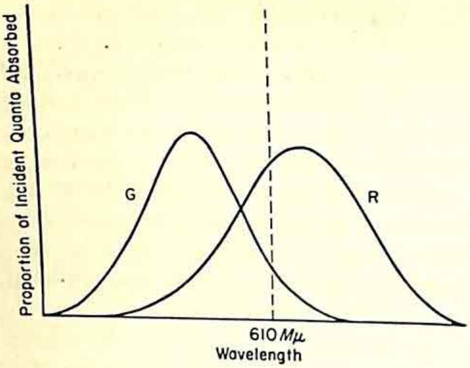


FIG. 6. Generalized spectral absorption curves for two hypothetical retinal pigments. (The wave length 610 mu, indicated on the abcissa, will be specifically referred to later in the text.)

cone types are shown in Figure 6.[3] The two cone systems will be labeled R and G, as in Figure 6. Now consider the situation in which a subject is shown a stimulus and is forced to call it either red or green, no other alternative being permitted. Let us assume that the probability of his giving a "red" judgment depends upon the ratio of the outputs of the R and G systems. In the simplest case, when the R output is greater, he will be more likely to call the stimulus red, when the G output is greater, he will call it green, and when they are equal, he will judge it green half the time and red the other half. There is no theoretical necessity that the output ratio be 1.0 for 50% "red" judgments, but that particular condition will be assumed for much of the remainder of this paper, merely to make the discussion easier to phrase.

Figure 7 shows the results obtained when a subject is forced to judge a flash of monochromatic light as either red or green, as the wave length in the region between red and green was varied by the method of constant stimuli. It should be noted that the slope of this curve is very steep, so that a change of just one or two millimu on either side of the 50% wave length will elicit reliable judgments of red or of green. Now if a subject is given a stimulus with a wave length right at the 50% point and asked what its hue *really* is, he says it is pure yellow. It follows from these observations that the last assumption discussed above is equivalent to assuming the following: when the R output exceeds the G output, the subject will see red, when the G output is greater, he will see green, and when the two are equal, he will see yellow. (This is of course true only for the spectral region between red and green. There may be other spectral regions in which the R and G outputs are also equal, but where the hue is given some other name, e.g., blue. The question of whether

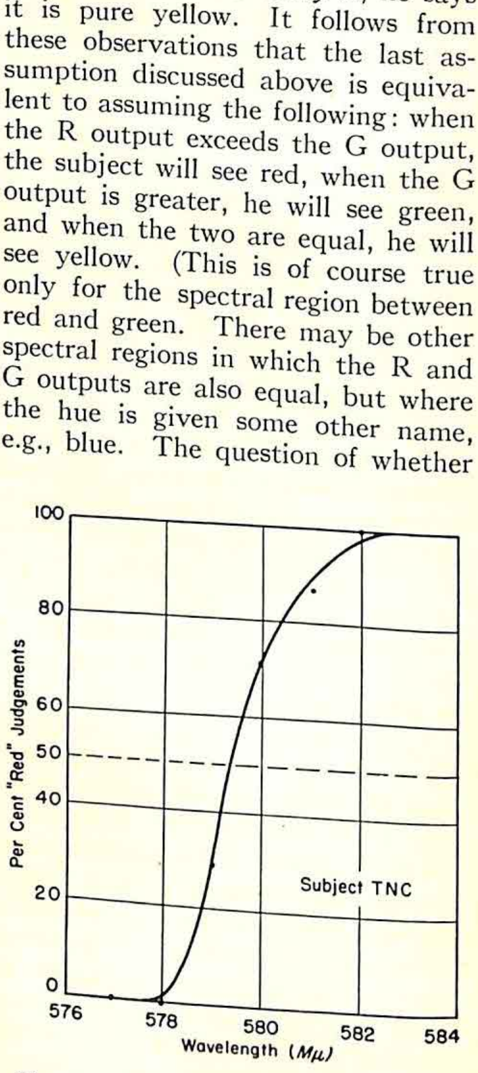


FIG. 7. Distribution of judgments for constant stimulus method. (On each 1/4-second flash, the subject was given a forced choice between responses "red" and "green.")

---

[3] The theory discussed in this paper requires that there be *at least* two cone systems, but does not approach the question of how many there may be altogether.

the yellow sensation is produced by a mixture of the R and G outputs or represents an excess of yellow over blue output with the red and green outputs cancelling each other, is not relevant to any of the present discussion.)

The explanation of some of the hue changes that occur to stimuli of high luminance follows directly from these assumptions. Monochromatic light in the long wave region of the spectrum, for example at 610 millimu, will be absorbed by both the R and the G cone pigments, but since the absorption coefficient for the R pigment is higher than the G at 610 millimu (see Figure 6), the R cones will absorb a greater percentage of the incident quanta. This is exactly equivalent to putting a greater illuminance on the R than the G cones. Therefore, if the R and G cones differ only in their absorption spectra, and a stimulus of very high luminance at 610 millimu is given, their output versus time curves will look like those in the lower part of Figure 2, Curve A representing the G and B the R system. The stimulus should look red at first, then change through yellow (when the curves cross) to green, and finally slowly change back through yellow to orange.

It has been reported in previous papers (Cornsweet et al., 1958) that a high-luminance red stimulus will appear to change from red through yellow to green. "No further color changes were observed, even though fixation was maintained for as long as three minutes" (p. 898). The present more extensive data at a variety of wave lengths, and at luminances greater than those used in the earlier study, clearly show that further changes do occur. In fact, the hue of the stimulus eventually changes back from green through yellow to orange as is predicted by the theoretical argument above. This later set of changes happens to take appreciably longer than 3 minutes at the relatively low luminances and at the particular wave lengths used in the earlier study (Cornsweet et al., 1958). But at other wave lengths and luminances, all of the changes appear to be over within a minute or two (see the data in succeeding sections).

GENERAL ASPECTS OF HUE CHANGES

The general course of the hue changes that occur when viewing a high-luminance stimulus at 620 millimicrons is illustrated in Figure 8. This figure is drawn from a set of crude binocular matchings, in which the high-luminance stimulus was presented to one eye and a matching stimulus of much lower luminance, whose wave length could be varied, was presented to the other. The curve in this figure should not be taken as anything more than a qualitative indication of the phenomenon under discussion. Quantitative data of a more reliable and useful sort will be presented later.

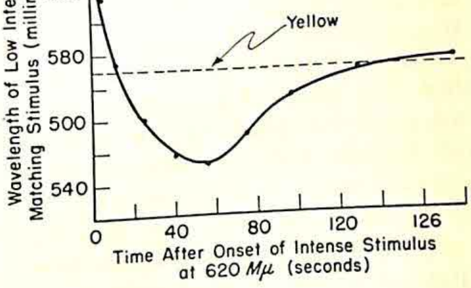


FIG. 8. Wave length of a low luminance stimulus flashed to the left eye that matches the hue of a high luminance stimulus, at 620 mu, steadily fixated by the right eye, as a function of the time of viewing the high-luminance stimulus.

At this writing, the limits of the range of wave lengths for which hue changes occur have not been well established. However, the following general statements are true. High-luminance stimuli with wave lengths in the "green-yellow" region of the spectrum (from about 560 to 580 millimu) look green at first, turn a much deeper green, then back through yellow to orange, and finally back to yellow or greenish yellow. Any wave length between 580 and 655 mu will look orange at first, change through yellow to a good, saturated green, and finally back to an orangish yellow. It is probable that light of wave lengths longer than 655 mu also appears to turn green if its luminance is high enough, but 655 mu is the limit with the present apparatus.

Stimuli below 560 mu have not been studied systematically, but in general they seem simply to desaturate to a white at the luminances available in the present apparatus.

For stimuli in the region of about 620 down to 560 mu or less, when the luminances are well above those necessary to produce the change to green, another change seems to be superimposed upon the first 15 or 20 seconds of viewing. The field begins its normal course of hue change, but then abruptly turns to a desaturated pink or violet. It retains that appearance for as long as 20 seconds, then changes back to orange, yellow, or green, and the normal sequence is resumed. This change to pink or violet has not yet been studied quantitatively.

All of the changes in appearance that have been observed depend in character and timing upon the luminance, and, of course, the wave length of the stimulus. These dependencies will be analyzed below.

QUANTITATIVE ANALYSIS OF HUE CHANGES

The photochemical theory outlined above requires that the time course of changes in hue be systematically related to the luminances and wave lengths of the stimuli. In order to evaluate the theory quantatively, data have been collected on the time course of hue changes as a function of retinal illuminance for a number of wave lengths. Those data and their relationship with the theory will be presented below.

PROCEDURE

Judgments Used for the Collection of Quantitative Data

The data presented here were collected by using two different sorts of procedures. The data for Figure 8 (and the data for the brightness reversals discussed earlier) are based upon a matching technique. Pairs of stimuli are found which differ physically but are judged the same along some perceptual dimension (hue in Figure 8, hue and brightness for the brightness reversal data). The other procedure used here is a sort of hue naming or identification. For example, most of the data in this report result from recording the times at which a subject judges the hue of the stimulus to be pure yellow, that is, neither reddish yellow nor greenish yellow.

In situations where a choice may be made between matching and hue naming techniques, it is usually far preferable to employ matching (Brindley, 1960). Hue naming was used principally here for two reasons. First, for the particular phenomenon studied, the problems introduced by matching are considerably worse than those of hue naming. For example, to determine the time at which a stimulus of high luminance produces a hue match with a standard hue (e.g., yellow) the matching field must be at a low enough luminance that its hue is not changing very much. Therefore, the luminance of the matching field should be at least 2 log units lower than of the test field. However, the stray light from the test field then approaches and perhaps even exceeds the direct illuminance of the matching field. If the matching field is presented to one eye and the test field to the other, the effects of the stray light are not eliminated, but just rendered less estimable. The collection of the data represented in Figure 8 took advantage of the fact that, even when the two fields are projected on to noncorresponding parts of the two retinas, rivalry occurs. That is, if the stimulus of high luminance is presented continuously and the matching field is occasionally flashed to the other eye, the only thing visible during the flash seems to be the matching field. It is then possible to find a test flash wave length that is the same hue as the high-luminance field at a given time after the onset of the high-luminance field. However, this procedure clearly involves so many factors of unknown effect (stray light, rivalry, transients from the flash, etc.) that it is at best useful only to give a rough idea of the sorts of hue changes that do occur.

A different sort of matching technique might have been used in this study. Instead of determining the time at which the intense field matches one of a fixed hue, the subject might have been presented with two patches having different wave lengths and high but different luminances. Then as the patches both changed in hue (and brightness), the time when they match could have been determined. While not all pairs of stimuli would reach a match, there must be sets of them that would do so. This kind of matching procedure would not be so severely affected by stray light, and in many ways is preferable to the color naming procedure. Brindley (1960) has clearly pointed out that this type of matching procedure (or at least one almost identical with it) enables one to draw physiological conclusions from psychophysical data, without requiring what he calls "psychophysical linking hypotheses." However, it is also true that, when such a procedure is used, psychophysical linking hypotheses *cannot* be tested, and one of the aims of this paper is to examine just such a question, namely, what are the physiological correlates of the condition under which a subject reports the hue as red half the time and green the other half. The matching technique however does have many advantages over color naming as long as psychophysical linking hypotheses are not in question, and studies using strict matching procedures are now in progress.

On the other hand, the procedure that consists of recording the instants in time when the subject calls the stimulus neither reddish nor greenish yields very reproducible results, and results which may be interpreted in a relatively straightforward way.

Apparatus

The apparatus used to collect all of the color data in the present study is shown schematically in Figure 9. The source, S, was a tungsten-coil-filament lamp (GE PH/18A/T10P), run from a constant voltage transformer at 6 volts, 18 amperes. Lens L_1 formed an image of the filament on the entrance slit, En, of a Bausch and Lomb grating monochromator of 500 millimeter focal length, with a 600 lines per millimeter grating, blazed in the visible. The optics of the monochromator imaged the entrance slit (and the filament) on the exit slit, Ex, with an equivalent aperture ratio of 4.4. An airspaced, coated, achromatic lens, L_2, focal length 150 millimeters, $f = 1.9$, imaged the exit slit in the plane of the pupil of the subject's eye so that he saw a Maxwellian view of the back surface of the lens L_2. The entrance and exit slits of the monochromator were set to widths of 1.5 millimeters, resulting in stimuli with spectral band widths of 5 millimu. The

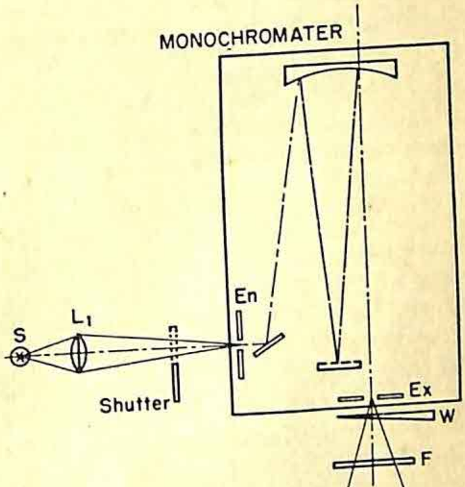


FIG. 9. Apparatus for delivering very intense monochromatic stimuli. (See text for description of specific elements.)

length of the exit slit was also set to 1.5 millimeters. Therefore, since Lens L_2 gave unit magnification, the image at the pupil was a square 1.5 millimeters on a side. This is small enough that as long as the exit slit is imaged at the center of the natural pupil, changes in the size of the natural pupil will not produce appreciable changes in retinal illuminance. A neutral-density wedge, W, and filters, F, were used to adjust the luminance of the stimuli. The wedge was close enough to the exit slit that no balancing wedge was needed. A clear piece of glass, G, reflected a small fraction of the stimulus light to a photomultiplier tube, PH, and the output of the phototube was used to monitor the luminance of the stimulus.[4]

A Hilger thermopile and mirror galvanometer were used to make the energy calibrations. The thermopile was placed at the pupil position during these measurements. The spectral purity of the stimulus lights was measured with a second monochromator set at a 5 millimu band width. When the stimulus was at any of the wave lengths studied, no other band had a luminance as great as 1% of the stimulus luminance, and the total light at other wave lengths was less than 5% of the stimulus luminance.

The subject's head was held in position by a dental-impression bar. At his hand was a control panel containing push buttons which operated timers, so that the timing of any changes in appearance could be recorded.

When the stimulus was on, the subject saw a circular field, uniformly illuminated, as illustrated in Figure 10. The dark circular line in this field was produced by cementing a circle of fine wire to the back surface of L_2. This circle helped the subject to make his judgments, in that he judged only the hue within the circle, ignoring the occasional edge effects that small eye movements produced at the outer margins of the field.[5]

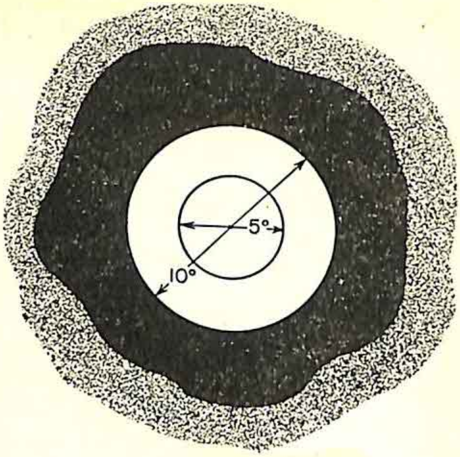


FIG. 10. Stimulus pattern as it appeared to the subject. (The entire 10-degree circular area was uniformly illuminated with monochromatic light. The subject was instructed to fixate the center of the pattern and judge the hue of the inner disk.)

RESULTS OF HUE CHANGE STUDY

The data plotted in Figure 11 show how the time taken to turn yellow varies as a function of the retinal irradiance of a monochromatic stimulus at 610 millimu. For example, at an irradiance of 3×10^{12} quanta per second per square millimeter of retina (uncorrected for ocular transmission losses), the patch looked red or orange at first. The subject pressed his key to indicate that it was yellow (neither reddish nor greenish) after 17 seconds of viewing. He pressed a second key to indicate the second occurrence of

---

[4] For the relatively short slit-length used here, the coil filament bulb delivered more energy through the system than did the ribbon filament bulb that is ordinarily supplied with the monochromater. However, using the coil filament, the energy at the exit slit varied somewhat from day to day in an unpredictable way. This variability was probably due to the fact that only a very small piece of the filament image was used, and, since the image is not uniformly intense, slight amounts of sagging of the coil or its supports could result in noticeable changes in output at the exit slit. For this reason, the monochromater output was continuously monitored, and the intensity settings were made from the photomultiplier output rather than from the wedge or filter calibrations.

[5] The particular choice of a 10-degree stimulus disc is not crucial. The hue changes occur with fields as small as 15 minutes of arc in diameter. (We have not tested still smaller sizes.) But the variance of the temporal aspects of the judgments is smallest when the field is large.

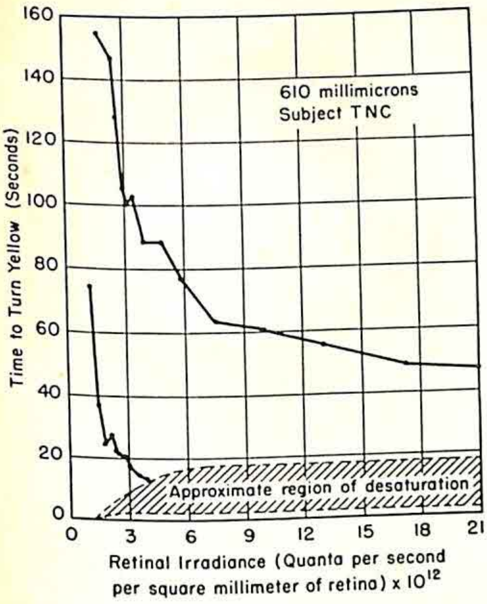


FIG. 11. Time to change hue as a function of the retinal irradiance of the stimulus, at 610 mu. (Each point is a mean of two readings.)

yellow 100 seconds after the stimulus had first been turned on. The area between the two curves represents the times during which the stimulus looked green.

The upper curve, for the second change to yellow, extends out to a retinal irradiance of 21 × 10^{12} quanta per second per square millimeter, the maximum irradiance available in the apparatus. The lower curve is not plotted beyond 3.9 × 10^{12} quanta/sec/mm^2 because the hue changes in that region were obscured by a change to desaturated pink or violet, as indicated roughly on Figure 11. For this wave length, the stimulus never appeared yellow at irradiances below 1.2 × 10^{12} quanta/sec/mm^2.

Figure 12 shows the same sort of data for light of three other wave lengths in the long wave end of the spectrum. The curves for 600 millimu are very similar to those for 610 millimu (Figure 11). As the wave length increases, the lower curves shift toward the higher irradiances, but retain the same general shape. The upper curve at 630 millimu shows evidence of a peak at the lower retinal irradiances. The upper curve at 640 millimu exhibits a very large amount of variability, and is not plotted beyond 9 × 10^{12} quanta/sec/mm^2 because the change back to yellow had not occurred during 210 seconds of continuous viewing. The variability in all cases increases very greatly as the time of viewing becomes extremely long. It will be shown that curves of this form may be expected from the theoretical considerations discussed previously.

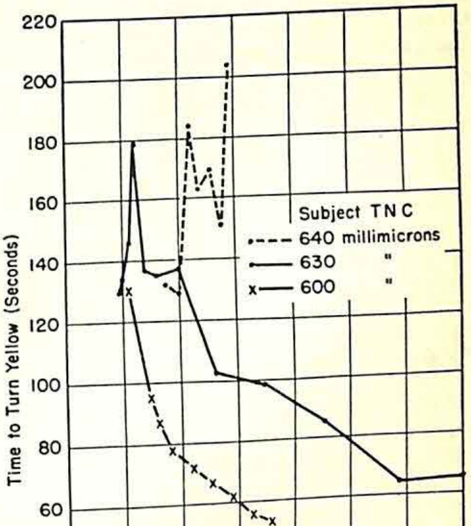


FIG. 12. Time to change hue as a function of retinal irradiance for three different wave lengths. (Each point is a single judgment.)

RELATIONSHIP BETWEEN THE DATA AND THE THEORY

The data in Figure 11 have been replotted as dots in Figure 13. They show how the time taken to turn yellow varies as a function of the irradiance of a monochromatic stimulus at 610 millimu. Every dot in this figure represents a time when the stimulus looked yellow. Therefore, according to the theoretical arguments outlined above, each of these points stands for a condition where the outputs of the R and the G systems are in some particular ratio. That is:

$$\Omega_r = C\Omega_g \qquad [8]$$

where C is the value of the ratio at which the hue is equally often called red and green in a forced choice. (In the preceding discussion C was assumed to be 1.0 for simplicity.)

Substituting the photochemical Equation 7 for output in Equation 8, and supplying the appropriate subscripts:

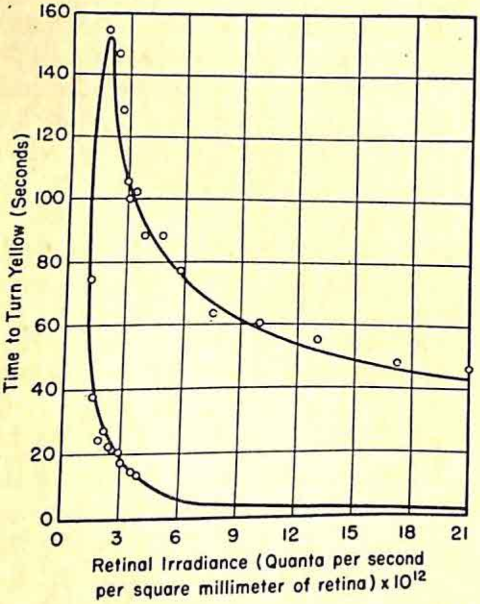


FIG. 13. Time to change hue as a function of retinal irradiance, for light at 610 mu. (Each point is a mean of two judgments. The smooth curve is a plot of the iterative solutions to the implicit Equation 10, using the parametric values listed in the text.)

$$A_rQM_r\left[\frac{K_r}{K_r+A_rQ} + \frac{A_rQ}{K_r+A_rQ}\cdot e^{-t(K_r+A_rQ)}\right] = CA_gQM_g\left[\frac{K_g}{K_g+A_gQ} + \frac{A_gQ}{K_g+A_gQ}\cdot e^{-t(K_g+A_gQ)}\right] \qquad [9]$$

where t is the time of occurrence of yellow. This equation may be rewritten:

$$\frac{K_rA_r}{K_r+A_rQ} + \frac{A_r^2Q}{K_r+A_rQ}\cdot e^{-t(K_r+A_rQ)} = J\left[\frac{K_gA_g}{K_g+A_gQ} + \frac{A_g^2Q}{K_g+A_gQ}\cdot e^{-t(K_g+A_gQ)}\right] \qquad [10]$$

where $J = M_gC/M_r$.

A crude iterative procedure has been used to find the values for the five parameters of this equation that yield a good fit to the data. The two smooth curves drawn in Figure 13 are a plot of Equation 10, when the following values of the parameters are used:

time constant of regeneration of red system, $K_r = .010 \text{ sec}^{-1}$

time constant of regeneration of green system, $K_g = .0083 \text{ sec}^{-1}$

photosensitivity of red system at 610 mu, $A_r = 1.70 \times 10^{-15}$

photosensitivity of green system at 610 mu, $A_g = 1.32 \times 10^{-15}$

ratio of amount of green to red pigment multiplied by the output ratio for yellow judgment, $J = 1.18$.

These preliminary values may be compared with Rushton's (1958) estimates:

$$K_r = .008 \text{ sec}^{-1}$$

$$K_g = .008 \text{ sec}^{-1}$$

$$A = .25 \times 10^{-15}$$

for the cone system as a whole, based upon measurements of the bleaching of cone pigments by white light, and corrected for an ocular transmission loss of 50%. Assuming that only an R and a G pigment are present, and correcting for 50% transmission losses, the estimate of A for the entire cone system on the basis of the present data is:

$$A_t = 1/.5A_{r,g} = 2[1-(1-A_r)(1-A_g)]$$

$$\approx 2(A_r + A_g)$$

$$A_t \approx 6 \times 10^{-15} \text{ cm}^2$$

The value of A_r determined from the color reversal data agrees much more closely with that quoted by Brindley (1960). From psychophysical data (quite different from those presented in the present paper), he estimates that $A_r = 1.61 \times 10^{-15}$ (p. 217). The writer has been unable to find estimates of the constant "J" in the literature. (This constant represents a psychophysical linking hypothesis, in Brindley's terminology.)

The agreement between these estimates is not perfect, nor is the fit in Figure 13. Its presentation is not intended to imply that the present formulation and these particular values are the "true" ones. The plot is meant to serve two purposes. First, it makes clear that the general formulation is consistent with the reported hue changes when put into at least this specific form. Second, it serves as a preliminary report on the extensive work now in progress attempting to evaluate, quantitatively, the specific theoretical arguments presented here. This work consists in making determinations of the best fit values of the parameters of Equation 10 under a number of independent conditions. For example, if this procedure is repeated at a number of different wave lengths in the red region of the spectrum, the best fit values of K_r, K_g, and J should be identical at each wave length, but the values of A_g and A_r should trace out the spectral sensitivity curves of the R and G systems.

The quantitative procedure used for finding parameters of the color phenomena may be applied to the achromatic brightness reversal phenomenon as well. However, the particular stimulus configuration used to gather the brightness reversal data presented in this paper do not lend themselves to straightforward analysis. For example, it is almost certain that the illuminances at the borders of a retinal region are more important in determining the apparent brightness of the region than are the illuminances in other regions; and, for a bar against a background, stray light and small eye movements have large and indeterminate effects on the effective illuminances at the borders. An experiment is currently in progress in which these factors are minimized, while the time course of brightness reversal is determined in much the same way as before. The results of this experiment should be directly analyzable in terms of the photochemical hypothesis presented above.

Observations on the rate at which the hue of the stimulus changes are qualitatively consistent with the theory, as plotted in the lower part of Figure 2. That is, the first change through yellow passes quickly, the stimulus actually appearing to be yellow for a very short time. The second change, where the theoretically determined output curves are running more nearly parallel, occurs very slowly and gradually. This is borne out quantitatively by the fact that the variance of the judgments of the time of second yellow point is appreciably greater than the variance of the first yellow point. The brightnesses of the stimuli also agree qualitatively with the theory. Although brightness has not been measured for the monochromatic stimuli, it is always high at first, gradually dropping to a much lower equilibrium level.

From the theory as it has so far been stated, it might be expected that, just as a red light turns green, a green light should turn red, and a yellow one should not change in hue. This would indeed be the prediction if it happened to be true that (*a*) the rates of regeneration of the red and green systems were identical, and (*b*) the density of red and green pigment molecules were also identical. But if one regeneration rate were greater than the other, high-luminance stimuli near the yellow region would, at *equilibrium*, take on the hue of the system that regenerates faster. From the present findings, this hue seems to be red, but the data are not yet complete enough to draw firm conclusions about the relative rates of regeneration.

The fact that yellow and greenish yellow stimuli change toward purer greens before changing back through yellow would follow directly from the theory if there is more green than red pigment present in the stimulated region of the retina. Consider, for example, light of a wave length that looks yellow when it is first turned on. This means that it is initially isomerizing red and green pigment molecules at the same rate (again assuming that $C = 1$ for convenience). However, if there are actually fewer red molecules present per unit of area, the rate of isomerization of red molecules will immediately become smaller than the corresponding rate for the green system, and the stimulus should turn green. This sort of effect is superimposed upon the processes already discussed so that, at the proper irradiances, an initially green stimulus should first get greener, then change back through yellow to red, and finally again through yellow to green, as it does.

It was mentioned above that, when the wave length is not too long (in the "orange") and the irradiance is extremely high, the early stages of hue change seem to be obscured by a wave of desaturated pink or violet. This effect would be expected if short wave sensitive systems were present and did produce desaturation. In this case, during the early stages of adaptation, both the R and the G systems are so intensely bombarded, relative to the short wave sensitive systems, that the R and G outputs drop to values comparable with the short wave systems, and desaturation occurs.

Note: One should exercise caution when viewing stimuli of very high luminance. Fechner permanently damaged his retinae by staring at the sun, and ophthalmologists can count on an increased patient load after solar eclipses. Since there have been very few systematic studies of the pathological effects of intense lights (see Whiteside, 1960), there are no clear rules on what should be avoided. In general, the sun should not be used as a source unless it is very strongly filtered, and any source that contains a large amount of infrared radiation should be filtered by heat-absorbing glass. The writer and many others who have observed the phenomena discussed in this paper exercised these precautions and experienced no discomfort except for afterimages that last for a few minutes.

SUMMARY

Evidence concerning the absorption of light quanta by visual pigment molecules is used to derive a theoretical relationship between the parameters of light incident on the retina and the output of the receptor cells. This relationship leads to the following prediction: when two retinal areas are illuminated, one very intensely and the other still more intensely, there will be a period of time during which the less intensely illuminated region will appear brighter than the more intensely illuminated one. Data verifying this prediction are presented.

When the theoretical treatment is expanded to include certain assumptions about retinal color mechanisms, it provides an explanation of some of the changes in hue that occur when a very intense colored stimulus is viewed steadily. Data on the temporal course of these hue changes are presented, and their relationship to the theory is discussed. This kind of analysis is shown to be a possible means for the quantitative evaluation of certain physiological properties of the visual system.

REFERENCES

AUERBACH, E., & WALD, G. The participation of different types of cones in human light and dark adaptation. *Amer. J. Ophthal.*, 1955, **39**, 24–40.

BRINDLEY, G. *Physiology of the retina and visual pathway.* London: Edward Arnold, 1960.

COHEN, J. Color adaptation to 1945. *Psychol. Bull.*, 1946, **43**, 121–140.

CORNSWEET, T. N., FOWLER, H., RABEDEAU, R. G., WHALEN, R. E., & WILLIAMS, D. R. Changes in the perceived color of very bright stimuli. *Science*, 1958, **128**, 898–899.

DEMOTT, D. W., & BOYNTON, R. M. Retinal distribution of entoptic stray light. *J. Opt. Soc. Amer.*, 1958, **48**, 13–21.

HECHT, S. A theory of visual intensity discrimination. *J. gen. Physiol.*, 1935, **18**, 767–789.

KROPF, A., & HUBBARD, R. The mechanism of bleaching rhodopsin. *Ann. N. Y. Acad. Sci.*, 1958, **74**, 266–280.

NODDACK, W., & JARZYK, G. Der Nachweis der Farbsehstoffe in lebenden menschlichen Auge. *Farbe*, 1955, **4**, 59–78.

RUSHTON, W. A. H. Kinetics of cone pigments measured objectively on the living human fovea. *Ann. N. Y. Acad. Sci.*, 1958, **74**, 291–304.

TROLAND, L. T. *Principles of psychophysiology.* Vol. 2. New York: Van Nostrand, 1930.

WHITESIDE, T. C. D. *Dazzle from nuclear weapons.* Washington, D. C.: National Academy of Science-National Research Council, 1960. (NAS-NRC Publ. No. 835)

(Early publication received
October 27, 1961)

Psychological Review
1962, Vol. 69, No. 4, 274–294

PRACTICE AS A PROCESS OF SIMPLIFICATION

MARSHALL B. JONES[1]

United States Naval School of Aviation Medicine, Pensacola, Florida

From time to time in the last 30 years a psychologist has intercorrelated successive trials in the practice of some task or other. And in every instance the correlations have arranged themselves in a very definite pattern.

The correlations which appear in Table 1 are a good example. The task here is the Two-Hand Coordination Test, which requires the subject to move a pin around a clover shaped runway by the coordinated turning of two control handles. The subjects are 152 basic airmen; and the trials are 60 seconds long, with 30-second rest intervals between trials.[2]

TABLE 1

INTERTRIAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.79	.77	.74	.73	.71	.71	.70
X_2			.87	.87	.84	.82	.82	.82
X_3				.91	.89	.87	.85	.86
X_4					.91	.88	.86	.88
X_5						.89	.90	.90
X_6							.93	.93
X_7								.94
X_8								

But notice the pattern. For each trial the correlations are highest in the superdiagonal (just above the main diagonal) and decrease monotonically as we move either upwards or to the right, until we come to the smallest correlation, which is in the upper right-hand corner of the matrix. There are a few inversions in the last two columns but they are minor, amounting in no case to more than .02.

This pattern, this "superdiagonal form" as I will call it, can be seen in every practice matrix which has ever been published. It seems to be a universal property of practice that a trial correlates strongest with its neighbors and weakest with the first and last trials in the series (Edgerton & Valentine, 1935; Fleishman, 1953, 1960; Fleishman & Hempel, 1954, 1955; Fleishman & Parker, 1959; Greene, 1943; Houston, 1950; Melton, 1947, p. 1022; Perl, 1934; Reynolds, 1952a, 1952b; Viteles, 1933).

PRINCIPAL POSSIBILITIES

Now what are we to make of this fact? Structurally, there are three principal possibilities. The first, in an historical sense, is the Woodrow two-factor structure.

In 1939 Herbert Woodrow observed that every practice matrix which was known at that time could be explained in terms of two common factors with positive weights which increased on one factor and decreased on the other as one went through the practice series. In other words, he posited that:

$$\begin{aligned}
X_1 &= a_1F_1 + b_1F_2 + c_1S_1\\
X_2 &= a_2F_1 + b_2F_2 + c_2S_2\\
X_3 &= a_3F_1 + b_3F_2 + c_3S_3\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
X_n &= a_nF_1 + b_nF_2 + c_nS_n
\end{aligned}$$

---

[1] Opinions or conclusions contained in this paper are those of the author. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

[2] For the use of these data I am indebted to E. A. Bilodeau.

where the X_i were the trial scores, F_1 and F_2 were the two common factors, the S_i were specific factors, and loadings, a_i and b_j, obeyed these laws:

$$0 \leq a_1 \leq a_2 \leq \cdots \leq a_{n-1} \leq a_n$$

and

$$b_1 \geq b_2 \geq b_3 \geq \cdots \geq b_n \geq 0$$

We now know that the Woodrow structure cannot explain all practice matrices. Sometimes, particularly in long matrices, a third Woodrow factor is required, or even a fourth. Nevertheless, the great bulk of practice matrices can be explained in terms of two Woodrow factors, and the present matrix is a case in point. If we hypothesize that:

$$\begin{array}{llll} a_1 = .61 & a_5 = .92 & b_1 = .57 & b_5 = .24 \\ a_2 = .75 & a_6 = .94 & b_2 = .57 & b_6 = .19 \\ a_3 = .83 & a_7 = .94 & b_3 = .47 & b_7 = .19 \\ a_4 = .87 & a_8 = .94 & b_4 = .39 & b_8 = .19 \end{array}$$

we get the correlations which appear in Table 2. Clearly, the fit is good. So, as far as the correlations are concerned, the Woodrow structure is a tenable hypothesis.

The second principal possibility is the "double accumulation structure." For an eight-variable matrix the equations are:

$$\begin{array}{l} X_1 = f_1+f_2+f_3+f_4+f_5+f_6+f_7 \qquad\qquad\qquad\qquad\quad +s_1 \\ X_2 = f_1+f_2+f_3+f_4+f_5+f_6+f_7 \qquad\qquad\qquad\quad f_{13}+s_2 \\ X_3 = f_1+f_2+f_3+f_4+f_5+f_6 \qquad\qquad\quad +f_{12}+f_{13}+s_3 \\ X_4 = f_1+f_2+f_3+f_4+f_5 \qquad\qquad +f_{11}+f_{12}+f_{13}+s_4 \\ X_5 = f_1+f_2+f_3+f_4 \qquad +f_{10}+f_{11}+f_{12}+f_{13}+s_5 \\ X_6 = f_1+f_2+f_3 \qquad +f_9+f_{10}+f_{11}+f_{12}+f_{13}+s_6 \\ X_7 = f_1+f_2 \qquad +f_8+f_9+f_{10}+f_{11}+f_{12}+f_{13}+s_7 \\ X^8 = f_1 \qquad\quad +f_8+f_9+f_{10}+f_{11}+f_{12}+f_{13}+s_8 \end{array}$$

where the trial scores are in standard form and all factor scores are in *deviation* form only.[3] In some respects double accumulation resembles Woodrow structure. Both hypotheses call for a two-process interpretation of practice, in which one process gains in importance with practice and the other loses. They differ, however, in the way these two processes are conceived. In the Woodrow structure each process is represented by a single factor with weights which decrease or increase as practice proceeds. In the double accumulation structure each process is represented by a set of factors in which the included factors accumulate or drop out with each new trial.

TABLE 2

THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST UNDER A WOODROW HYPOTHESIS

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.78	.77	.75	.70	.68	.68	.68
X_2			.89	.87	.83	.81	.81	.81
X_3				.91	.88	.87	.87	.87
X_4					.89	.89	.89	.89
X_5						.91	.91	.91
X_6							.92	.92
X_7								.92
X_8								

The double accumulation structure is a special case of the "open contiguity structure," and there are some practice matrices which require the more general form. In the present example, however, if we hypothesize that $\sigma_{f_1}{}^2 = .70$ and that:

$$\begin{array}{llll} \sigma_{f_2}{}^2 = .00 & \sigma_{f_5}{}^2 = .02 & \sigma_{f_8}{}^2 = .01 & \sigma_{f_{11}}{}^2 = .02 \\ \sigma_{f_3}{}^2 = .00 & \sigma_{f_6}{}^2 = .02 & \sigma_{f_9}{}^2 = .03 & \sigma_{f_{12}}{}^2 = .04 \\ \sigma_{f_4}{}^2 = .03 & \sigma_{f_7}{}^2 = .02 & \sigma_{f_{10}}{}^2 = .02 & \sigma_{f_{13}}{}^2 = .12 \end{array}$$

---

[3] In the double accumulation structure and the simplex the right-hand side of each structural equation should be multiplied by an arbitrary constant. Otherwise, the structure conditions the variances and covariances as well as the correlations. Ordinarily, these constants are dropped as their inclusion would only complicate the equations to no particular advantage. Whenever we deal with variances or covariances, however, these constants must be reinstated.

TABLE 3

THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST UNDER A DOUBLE-ACCUMULATION HYPOTHESIS

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.79	.77	.75	.73	.70	.70	.70
X_2			.89	.87	.85	.82	.82	.82
X_3				.91	.89	.86	.86	.86
X_4					.91	.88	.88	.88
X_5						.90	.90	.90
X_6							.93	.93
X_7								.94
X_8								

the double accumulation structure generates the correlations which appear in Table 3. In this instance, therefore, as in most practice matrices, the correlations are easily explained in terms of a double accumulation hypothesis.

The third principal possibility, the simplex, differs fundamentally from the first two in that it calls for a one-process interpretation of practice. Practice, under a simplicial interpretation, is a process of simplification, exclusively. For an eight-variable matrix the equations are:

$$\begin{aligned}
x_1 &= f_1+f_2+f_3+f_4+f_5+f_6+f_7+f_8+s_1\\
x_2 &= \phantom{f_1+{}} f_2+f_3+f_4+f_5+f_6+f_7+f_8+s_2\\
x_3 &= \phantom{f_1+f_2+{}} f_3+f_4+f_5+f_6+f_7+f_8+s_3\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
x_7 &= \phantom{f_1+f_2+f_3+f_4+f_5+f_6+{}} f_7+f_8+s_7\\
x_8 &= \phantom{f_1+f_2+f_3+f_4+f_5+f_6+f_7+{}} f_8+s_8
\end{aligned}$$

where *all* variables are in deviation form only.[4]

To generate specific correlations from a simplex we must hypothesize two kinds of parameters: the communalities

$$h_i^2 = \frac{\sigma_{f_i}{}^2 + \cdots + \sigma_{f_n}{}^2}{\sigma_{x_i}{}^2}$$

and the complexities

$$c_i = \left(\frac{\sigma_{f_i}{}^2 + \cdots + \sigma_{f_n}{}^2}{\sigma_{f_i}{}^2 + \cdots + \sigma_{f_n}{}^2}\right)^{\frac{1}{2}}$$

In the present instance, if we set:

$h_1^2=.79$ $h_5^2=.91$ $c_1=1.00$ $c_5=.86$
$h_2^2=.87$ $h_6^2=.93$ $c_2=.93$ $c_6=.83$
$h_3^2=.91$ $h_7^2=.94$ $c_3=.90$ $c_7=.83$
$h_4^2=.91$ $h_8^2=.94$ $c_4=.88$ $c_8=.83$

we get the correlations which appear in Table 4 and a good fit. This particular example, moreover, is not at all exceptional. Every known matrix of correlations between trials of practice can be fit simplicially.

I should also point out that, right now, there does not exist any way of testing the significance of these fits. However, no one of these hypotheses leaves a residual correlation greater than .02, and we hardly need a test to tell us that residuals of this magnitude are not significant. Even if they were significant, the fact would be of no particular moment. True, the hypotheses would have to be modified;

TABLE 4

THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST UNDER A SIMPLICIAL HYPOTHESIS

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.77	.77	.75	.73	.71	.71	.71
X_2			.86	.85	.83	.81	.81	.81
X_3				.89	.87	.86	.86	.86
X_4					.89	.87	.87	.87
X_5						.89	.89	.89
X_6							.93	.93
X_7								.94
X_8								

---

[4] The requirement in the simplex that all variables be in deviation form only is important. Were we to require that the trial scores be expressed in standard form or, which is the same thing, that the trial scores all have the same variance (except for the arbitrary constants), then the "simplex" would become a single accumulation structure —which is a perfectly valid structure—but it has no place in the context of practice because it does not generate a superdiagonal form. For more on the single accumulation structure, see Jones (1960, pp. 50–51).

but the modifications would be so paltry in amount that nothing of any real consequence would be altered.

MOLAR CORRELATIONAL ANALYSIS

To the learning theorist the next question is obvious: Which of these three structures best expresses the true nature of practice? I intend to answer this question, but first I want to say a word or two about correlational analysis.

For the last 30 years the treatment of correlation matrices in psychology has been completely dominated by factor analysis. There has been criticism, of course. In fact, there are a great many psychologists, perhaps even a majority, who won't have any part of factor analysis. But for two generations now there has been no alternative to factor analysis. If a man calculated a correlation matrix, the only existing way of summarizing his result in structural terms has been the factor analytic way. In the last few years, however, this situation has changed. Beginning with Louis Guttman's work on the simplex in 1954 and continuing with my two monographs on correlational analysis (Jones, 1959, 1960), an entirely new discipline has been created, a completely different way of treating a correlation matrix.

This new discipline, molar correlational analysis, is coextensive with factor analysis in practice and absolutely incompatible with it in theory. The essence of this incompatibility is a matter of approach. Factor analysis is an inductive method. The factor analyst starts with the correlations and then attempts by a perfectly mechanical computing regime to grind out his factor structure. He knows very well, of course, that there are a great many, substantially different structures any one of which could generate the correlations. But the factor analyst has faith. Aside from their intercorrelations he may know nothing whatever about the variables which enter into the matrix. Nevertheless, he remains quite confident that by estimating communalities, factoring the matrix, and rotating axes (according to his preferred method) he will arrive at the structure which is actually responsible for the correlations being what they are.

The molar analyst is a much less believing creature. As he sees it, correlational analysis is a strictly deductive discipline. If we are given some hypothesis as to how the variables are structured, then algebra alone can tell us what the correlations ought to be—but that's all it can do. The construction of theories is a business for the psychologist, as psychologist, not the statistician. All the statistician can do is to give us the means for tracing out the consequences of our thinking. He can tell us which patterns follow from which structures. He can catalogue the more important of these structures and these patterns. He can even show us how to fit a structure to a correlation matrix. But beyond this he simply cannot go.

In itself, this difference would enable us to construct an entirely different approach to the treatment of correlation matrices. But there are several other respects in which molar analysis differs fundamentally from the factor analytic tradition. The most important of these is the emphasis which the molar analyst places upon correlation pattern. A correlation matrix is more than a collection of correlations. As often as not, these correlations, or a substantial number of them, display a definite pattern. We have already seen the superdi-

SIMPLICIAL ARGUMENT

Let's return now to the problem at hand, the superdiagonal forms of practice and their structural interpretation. At the present stage of the game the arguments for and against our three main hypotheses are far from decisive. In my opinion, the simplex has the best of it on the available evidence, but we have a long way to go before any one of these hypotheses or, for that matter, any other explanation of the correlations can be regarded as definitely established.

The case for the simplex rests on two main arguments. The first, and most weighty, comes from experimental psychology. Since the days of Bekhterev, psychologists have known that learning involved the progressive focussing of behavior on the one response which was successful in securing reward. At first the dog starts wildly in the stock. He makes all sorts of inappropriate responses; indeed, initially, it is only by accident that he lifts his paw. As practice continues, however, his behavior gradually focalizes, until in the end the dog barely lifts his paw the amount required to avoid the shock. As we observe it, in other words, practice is a process of simplification. How very natural therefore, that in its differential composition (which we do not observe) practice should also become simpler and simpler as one trial follows another.

In making this argument I want to be quite clear that I am talking about practice, not learning. There are some forms of learning, problem solving for example, which almost certainly are not processes of simplification. If Krechevsky (1932) had intercorrelated successive trials in his famous studies of "hypothesis" formation in rats, it is quite possible that the correlations would have shown the superdiagonal form. But in this case the originating structure could hardly have been a simplex. The rats' "hypotheses," after all, came and went. First they tried one thing and then another, first a spatial hypothesis and then a visual one. This is the kind of behavior which generates an open contiguum, a structure in which factors can come as well as go, not a simplex, in which they can only go (or come).[7] By "practice" I mean that portion of the learning sequence which follows the formation of the correct hypothesis. Learning doesn't end just because the person or animal has finally understood the problem. There remains the "stamping-in," the mastery of the problem. And it is this stamping-in, this practice, which the simplicial hypothesis requires to be a process of simplification.

The second argument has to do with the external relations of early and late stages of practice. According to the simplicial hypothesis, the early stages of practice contain a variety of factors in addition to those which are most particularly required in the practice task itself. This hypothesis implies that, if we correlate the trials of practice with materials which are categorically distinct from the practice task, then, on the average, the early stages of practice should be more strongly related to the test materials than the later ones. On the other hand, if we correlate the trials of practice with very similar materials, we should expect just the reverse result: in general, the later stages should be more strongly related to the test materials than the early ones.

---

[7] If a correlation matrix can be explained by a simplex with progressive simplification, it can also, as a matter of analytic necessity, be explained by a simplex with progressive complications. From a correlational point of view alone the possibility cannot be dismissed that the trials become increasingly complex with practice. I have rejected this notion for the same reasons I have accepted its inverse of progressive simplification, namely, its contradiction by the overt behavior of the subjects and by the correlations of the trials of practice with external measures.

Unfortunately, the only evidence we have which is relevant to these two consequences concerns psychomotor tasks. In this one area, however, the evidence is very clear-cut. When they used printed tests as the external measures, Adams (1953, 1957), Fleishman (1960), Fleishman and Hempel (1954, 1955), and Reynolds (1952a, 1952b) all found that the predictability of the trials on a psychomotor task generally decreased with practice. On the other hand, these same investigators have also shown that the relationship between one psychomotor task and another usually increases with practice, on either one of the two tasks or both. And if we take an area more or less in between, say training criteria, we find a more or less neutral result: for example, the correlation between the Air Force pass-fail pilot criterion and scores on the Complex Coordination Test neither increases or decreases with practice. Unless we make special assumptions, however, none of these results follows from either the Woodrow or the double accumulation structure. All told, therefore, the existing evidence on the relationship between trials of practice and external measures strongly supports the simplicial hypothesis.

IDENTIFYING FACTORS

In all branches of study the recognition of form is followed by the analysis into parts. Once the astronomer has recognized a star he wants to know what it's made of. Once the biochemist has isolated a substance he wants to know it's composition. And once the psychologist has decided, even tentatively, that practice is a process of simplification he wants to know what the elements are which participate in this process; he wants to identify the simplicial factors.

Now, the only way the psychologist can identify these f_i is to administer a variety of other tests to the same subjects who worked on the practice task. Once this is done, he can correlate the reference tests with the stages of practice and, through these correlations, obtain the correlations between the reference tests and the simplicial factors. If these final correlations are high, then the factors are identified. If not, we must try again, with other tests.

To illustrate this logic let us rework the results of a recent study by E. A. Fleishman.[8] In Table 5 I have reproduced the intertrial correlations on

---

[8] In his analysis of these data Fleishman threw both the trials of practice and the reference tests into one grand matrix and performed a factor analysis. There is, I think, a fatal objection to this procedure, quite apart from Fleishman's use of factor analysis. It is that the structure of the *practice task* becomes dependent upon the test materials. If we use one battery of tests, we will get one structure for the trials of practice; if we use another, we will get a second; if we use a third battery, we will get still another. But it's obvious, it seems to me, that the trials of practice on, say, the Rotary Pursuit Test have *some* structure, altogether apart from the tests we happen to administer. This structure doesn't have to be a simplex. As a factor analyst, Fleishman might very well prefer the Woodrow structure. But however we interpret them, it seems quite clear that there is only one true structure for a single set of variables. Under Fleishman's approach the factors we are trying to identify are a function of the materials with which we are trying to identify them. The only way to resolve this dilemma is the molar way, to analyze the intertrial correlations first, apart from the reference tests, and then, having arrived at a structural recognition of the factors, to attempt an indentification in substance.

TABLE 5

INTERTRIAL CORRELATIONS FOR THE ROTARY PURSUIT TEST

Trial[a]	1	3	5	7	9	11	13	15
1		.82	.73	.72	.65	.65	.58	.54
2			.88	.84	.77	.74	.70	.64
5				.90	.86	.83	.79	.74
7					.89	.86	.83	.80
9						.88	.87	.82
11							.88	.84
13								.85
15								

[a] Fleishman reported his results for the odd-numbered trials only.

TABLE 6

THEORETICAL CORRELATIONS FOR THE ROTARY PURSUIT TEST UNDER A SIMPLICIAL HYPOTHESIS

Trial	1	3	5	7	9	11	13	15
1		.79	.74	.72	.66	.63	.61	.56
3			.86	.83	.77	.73	.71	.65
5				.91	.85	.80	.77	.72
7					.91	.86	.83	.77
9						.91	.89	.82
11							.91	.84
13								.84
15								

the practice task, which was the Rotary Pursuit Test. This matrix has the superdiagonal form. And if we hypothesize a simplex and set:

$h_1^2 = .86 \quad h_5^2 = .96 \quad c_1 = 1.00 \quad c_5 = .73$

$h_2^2 = .92 \quad h_6^2 = .94 \quad c_2 = .89 \quad c_6 = .70$

$h_3^2 = .94 \quad h_7^2 = .90 \quad c_3 = .82 \quad c_7 = .69$

$h_4^2 = .98 \quad h_8^2 = .82 \quad c_4 = .78 \quad c_8 = .67$

we get the theoretical correlations which appear in Table 6. Since the fit is good, our hypothesis is tenable. It remains to find out what the hypothesized factors are, not structurally (we already know that) but substantially.

In addition to the Rotary Pursuit Test Fleishman administered a battery of 17 reference tests, 7 paper-and-pencil tests, and 10 apparatus tests.[9] Now,

TABLE 7

CORRELATIONS BETWEEN THE REFERENCE TESTS AND THE TRIALS OF PRACTICE

Test	Trials							
	1	3	5	7	9	11	13	15
General mechanics	.19	.22	.22	.18	.20	.19	.15	.12
Tool functions	.19	.22	.18	.16	.17	.14	.13	.07
Speed of identification	.20	.24	.26	.26	.29	.26	.27	.26
Instrument comprehension	.20	.22	.22	.20	.21	.25	.20	.23
Visual pursuit	.30	.33	.25	.23	.23	.24	.22	.21
Aiming	.27	.32	.34	.29	.32	.30	.27	.29
Marking accuracy	.26	.33	.33	.31	.34	.35	.38	.34
Purdue assembly board	.20	.21	.18	.16	.17	.15	.17	.17
Santa Ana dexterity	.28	.36	.39	.38	.32	.36	.37	.34
Rate control	.25	.25	.19	.21	.18	.17	.12	.19
Linear pursuitmeter	.23	.22	.18	.21	.20	.17	.18	.18
Rate of movement	.19	.22	.23	.21	.23	.21	.18	.26
Rotary aiming	.27	.25	.29	.27	.29	.26	.29	.29
Reaction Time I	.24	.24	.26	.24	.22	.19	.18	.25
Track tracing	.34	.34	.35	.34	.32	.32	.32	.33
Complex coordination	.37	.36	.34	.35	.32	.31	.35	.39
Reaction Time II	.21	.29	.32	.30	.27	.28	.32	.34

---

[9] The reader will notice that for the printed tests, Tests 1 through 7, the early trials are more strongly related than the late ones,

TABLE 8

CORRELATIONS BETWEEN THE REFERENCE TESTS AND THE SIMPLICIAL FACTORS

Test	Factors							
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
General mechanics	.02	−.03	.00	.06	−.11	.00	.11	.21
Tool functions	−.07	.08	−.10	.03	−.04	−.21	.11	.21
Speed of identification	.05	.08	.00	.12	.00	.11	.21	.22
Instrument comprehension	.14	−.11	.20	.06	−.07	.00	.11	.22
Visual pursuit	.07	−.03	.07	.06	−.04	−.42	.11	.33
Aiming	.14	−.03	.03	.12	−.15	.11	.21	.30
Marking accuracy	.02	.19	.13	1.2	−.04	.00	.32	.29
Purdue assembly	.02	.05	.00	.06	−.07	−.16	.05	.22
Santa Ana dexterity	.02	.14	.23	−.12	.00	.16	.37	.31
Rate control	.19	−.05	.00	−.06	.11	−.32	.00	.27
Linear pursuitmeter	.05	.05	−.07	.06	.11	−.21	−.05	.25
Rate of movement	.28	.03	−.03	.06	−.04	.05	.11	.21
Rotary aiming	.09	.11	.00	.09	−.04	.21	−.16	.30
Reaction Time I	.26	.03	−.07	−.03	−.04	.11	.00	.26
Track tracing	.12	.05	.07	.00	.00	.05	.00	.37
Complex coordination	.21	.22	.00	−.03	.07	−.11	−.11	.41
Reaction Time II	.16	.19	.13	−.06	−.07	.16	.42	.23

if we assume that the test-trial correlations, which appear in Table 7, are mediated by the simplicial factors only, we can calculate the correlations between the tests and the simplicial factors (Jones, 1960, pp. 28–31, 80–83). These last correlations, of course, are not necessarily exact, because the assumption under which they were calculated could be in error. However, when the simplicial communalities are as high as they are for the Rotary Pursuit Test, ranging from .86 to .98—and the test-trial correlations are as low as they are in this instance—we run little risk of serious error in accepting the hypothesis of exclusive simplicial mediation. And under this hypothesis, the correlations between the reference tests and the simplicial factors take on the values which appear in Table 8.

These values, unhappily, are low. Not one of them ranges beyond the .40s and only three are that high. In this particular instance, therefore, our effort to identify the simplicial factors was a failure. Nevertheless, there is nothing the matter with the method; if we persist with it, then sooner or later we will find tests which correlate strongly with the simplicial factors. When this happens, we will have identified the f_i.

CONCLUSION

In the end, of course, hypothesizing structures and identifying their factorial elements will not suffice. If we really want to understand the superdiagonal forms of practice, there is nothing for it but to experiment with them. How would these patterns change as a function of distribution of practice, or of intermittent reward? How would they be affected by prior practice on a related task? Sooner or later we must augment the passive

---

whereas for the apparatus tests, Tests 8 through 17, just the reverse is true. In this case, these tendencies are weaker than they are in any other published study, but even so they are still there.

arts of correlational analysis with active experimentation. We must theorize, of course; and this theorizing will inevitably be correlational in character. But we must also experiment.

REFERENCES

ADAMS, J. A. The prediction of performance at advanced stages of training on a complex psychomotor task. *USAF Hum. Resour. Res. Cent. Res. Bull.*, 1953, No. 53–49.

ADAMS, J. A. The relationship between certain measures of ability and the acquisition of a psychomotor criterion response. *J. gen. Psychol.*, 1957, **56**, 121–134.

EDGERTON, H. A., & VALENTINE, W. L. A. A factor analysis of learning data. *Psychol. Bull.*, 1935, **32**, 719.

FLEISCHMAN, E. A. A factor analysis of intra-task performance on two psychomotor tests. *Psychometrika*, 1953, **18**, 45–55.

FLEISHMAN, E. A. Abilities at different stages of practice in rotary pursuit performance. *J. exp. Psychol.*, 1960, **60**, 162–171.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. Changes in factor structure of a complex psychomotor task as a function of practice. *Psychometrika*, 1954, **19**, 239–252.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49**, 301–312.

FLEISHMAN, E. A., & PARKER, J. F., JR. Prediction of advanced levels of proficiency in a complex tracking task. *USAF WADC tech. Rep.*, 1959, No. 59–255.

GREENE, E. B. An analysis of random and systematic changes with practice. *Psychometrika*, 1943, **8**, 37–53.

GUTTMAN, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld, (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press, 1954.

HOUSTON, R. C. An evaluation of the predictive properties of measures of variability of performance on three psychomotor tasks. Unpublished PhD thesis, University of Maryland, 1950.

JONES, M. B. Simplex theory. *USN Sch. Aviat. Med. Monogr.*, 1959, No. 3.

JONES, M. B. Molar correlational analysis. *USN Sch. Aviat. Med. Monogr.*, 1960, No. 4.

KRECHEVSKY, I. "Hypotheses" vs. "chance" in the pre-solution period in sensory discrimination. *U. Calif. Publ. Psychol.*, 1932, **6**, 27–44.

MELTON, A. W. (Ed.) *Apparatus tests.* Washington: United States Government Printing Office, 1947.

PERL, R. E. An application of Thurstone's method of factor analysis to practice series. *J. gen. Psychol.*, 1934, **11**, 209–212.

REYNOLDS, B. Correlations between two psychomotor tasks as a function of distribution of practice on the first. *J. exp. Psychol.*, 1952, **43**, 314–348. (a)

REYNOLDS, B. The effect of learning of the predictability of psychomotor performance. *J. exp. Psychol.*, 1952, **44**, 189–198. (b)

VITELES, M. S. The influence of training on motor test performance. *J. exp. Psychol.*, 1933, **16**, 556–564.

(Received October 19, 1960)

SUPPLEMENTARY NOTE

In this note I propose to take up the fitting procedures for the three structures which were treated in the text of this paper, the Woodrow two-factor structure, the double accumulation structure, and the simplex. As a necessary preliminary, I will also treat a few, more elementary forms.

Mathematically, the best way to fit any molar structure is by the method of least squares. Given a structural hypothesis, we know that the theoretical correlations, ρ_{ij}, are some function of the structural parameters, $p_1, p_2, \cdots, p_n$. We could, therefore, proceed to minimize the squares of the differences between the theoretical and the empirical correlations, i.e., between the ρ_{ij} and the r_{ij}. To do this we need only differentiate the expression:

$$\sum_i \sum_j (\rho_{ij} - r_{ij})^2$$

and solve the normal equations which result.

There is, however, a difficulty with this approach: The number of normal equations which must be solved in order to determine a best fit is prohibitive. If we use least squares, every fitting process requires us to solve at least as many simultaneous equations as there are variables in the matrix. And this, in its turn, requires computing equipment of a high order. It is desirable, therefore, to find some other, less heroic means of arriving at essentially the same results.

In all the fitting procedures which I will describe the general process is the same. We begin by positing a tentative set of parameters, which we then adjust until we get what appears to be a best fit. These tentative values, of course, are not taken at random. Knowing what the structure is, we can determine tentative values which will require a minimum of adjustment. The only problem is to develop sound procedures for arriving at these tentative values.

SPEARMAN UNIT FACTOR

Let us begin with the Spearman unit factor, i.e., with the hypothesis that the variables which enter into the matrix contain a single common factor. If the matrix is arranged so that the variable with the biggest loading on this one factor comes first, the variable with the next biggest loading second, and so on, down to the variable with the smallest loading, which comes last, the structure generates a characteristic pattern. This pattern, which I call a "monotonic hierarchy," locates the biggest correlation in the upper left-hand corner of the matrix and the smallest in the lower right-hand corner with the others spread out in between. If the loadings are arranged the other way around, i.e., with the smallest loading first and the biggest last, then the pattern will be reversed: The biggest correlation will be in the lower right-hand corner and the smallest in the upper left-hand corner.

TABLE A1

A MONOTONIC HIERARCHY

Trial	X_1	X_2	X_3	X_4
X_1		.63	.64	.66
X_2			.70	.72
X_3				.74
X_4				

TABLE A2

SOME THEORETICAL CORRELATIONS UNDER A SPEARMAN UNIT–FACTOR INTERPRETATION

Trial	X_1	X_2	X_3	X_4
X_1		.63	.64	.66
X_2			.70	.72
X_3				.73
X_4				

The matrix in Table A1 displays this second form of the monotonic hierarchy. On a purely qualitative basis, therefore, the hypothesis of a Spearman unit factor is tenable. But before we can accept this hypothesis as a possible explanation of the correlations, we must first specify values for the factor loadings and attempt a fit.

Under a Spearman hypothesis the empirical correlations:

$$r_{ij} \approx a_i a_j$$

where a_i is the loading on the one common factor of the ith variable in the matrix. It follows that:

$$a_1 \approx \left(\frac{r_{12} r_{13}}{r_{23}} \right)^{\frac{1}{2}}$$

$$a_1 \approx \left(\frac{r_{1(n-1)} r_{1n}}{r_{(n-1)n}} \right)^{\frac{1}{2}}$$

In other words, working with the two "ends" of the matrix we obtain two tentative values for the first factor loading, a_1. If we compromise these two values, we will get a "balanced" single value, which, in this instance, equals .76. Then, since:

$$a_j \approx \frac{r_{ij}}{a_1}$$

we can go ahead and determine the tentative values for the remaining variables. In the case at hand we get these values:

$$a_1 = .76 \qquad a_3 = .84$$
$$a_2 = .83 \qquad a_4 = .87$$

Generating the theoretical correlations from these parameters we get the matrix which appears in Table A2. Ordinarily, we are able to improve upon the fit by adjusting the a_i, but in this instance the fit is just about perfect as it stands, since the biggest residual correlation is .01, and only one is that big. It's quite clear, therefore, that the matrix can be explained as originating in a single common factor.

AUGMENTED SPEARMAN UNIT FACTOR

By an augmented Spearman unit factor I mean a two-factor structure in which the loadings of all the variables on one of the two factors are equal. Correlationally, the effect of this second factor is to add a constant quantity, which equals the loading on this second factor squared, to all of the correlations in the matrix. Now, consider the matrix which appears in Table A3. Qualitatively, the pattern it displays, i.e., the monotonic hierarchy, is consistent with an augmented-Spearman interpretation. It remains to be seen, however, whether or not a good quantitative fit is possible.

In fitting an augmented Spearman hypothesis our first task is to determine the magnitude of the one loading on the second factor. Consider, therefore, the three "principal points" of the matrix, i.e., the upper left-hand corner, the upper right-hand corner, and the lower right-hand corner. Under a *nonaugmented* Spearman hypothesis:

$$r_{12}r_{(n-1)n} - r_{1n}r_{2(n-1)} \approx 0$$

TABLE A3

ANOTHER MONOTONIC HIERACHY

Trial	X_1	X_2	X_3	X_4
X_1		.88	.77	.71
X_2			.73	.67
X_3				.63
X_4				

TABLE A4

SOME THEORETICAL CORRELATIONS UNDER AN AUGMENTED SPEARMAN UNIT-FACTOR INTERPRETATION

Trial	X_1	X_2	X_3	X_4
X_1		.88	.77	.71
X_2			.74	.69
X_3				.63
X_4				

If this quantity is either negative or zero, then the fit cannot be improved by means of the additive constant (which must be positive). If this difference is positive, however, as in the present instance, then by subtracting a fixed positive amount from all the correlations in the matrix, the theoretical correlations can be brought more nearly into line with their empirical counterparts. Now, if we subtract .10 from all the correlations in the matrix, the tetrad difference between the principal points remains positive. In fact, it remains positive even if we subtract .50 from all of the correlations. If we subtract .60, however, the difference becomes negative. Suppose, therefore, that we take .50 as our tentative value for b^2, the one loading on the second factor squared. If we subtract this value from all of the correlations in the matrix and fit a Spearman unit factor to the residual correlations, in the manner already described, we get these values for our five parameters:

$$a_1 = .66 \quad a_3 = .41 \quad b = .71$$
$$a_2 = .58 \quad a_4 = .32$$

which generate the correlations which appear in Table A4. Once again, there is no need to adjust the tentative values for the parameters. But again it should be understood that adjustments are ordinarily required.

HALF-WOODROW TWO-FACTOR STRUCTURE

In the half-Woodrow, two-factor structure the loadings on one of the two factors are the same for the first half of the series, while on the other factor the loadings are the same for the second half of the series. In an eight-variable structure, for example, the loadings on the

TABLE A5

A THREE-CORNERED PATTERN

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.88	.77	.71	.69	.70	.72	.74
X_2			.73	.67	.65	.67	.69	.70
X_3				.63	.62	.64	.65	.67
X_4					.60	.61	.63	.64
X_5						.63	.64	.66
X_6							.70	.72
X_7								.74
X_8								

first factor:

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$$

and the loadings on the second factor:

$$\beta_5 = \beta_6 = \beta_7 = \beta_8$$

For the rest of it:

$$\alpha_5 \leq \alpha_6 \leq \alpha_7 \leq \alpha_8$$

and

$$\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$$

just as in the full-Woodrow structure.

Correlationally, however, the two structures are very different. The full-Woodrow generates a superdiagonal form, while the half-Woodrow generates a three-cornered pattern. Table A5 presents an eight-variable example. The correlations are high in the three corners of the matrix and drop off as one moves from any one of them toward the middle of the last column, the middle of the first row, or the middle of the superdiagonal.

A three-cornered pattern is naturally divided into three submatrices: the two apical triangles, which contain the correlations *within* the two halves of the series, and the central block, which contains the correlations *between* them. In the two apical triangles the correlations are generated by augmented Spearman hypotheses. So, we already know how to fit these correlations. In fact, we already have. The two apical triangles of Table A5 are the same two monotonic hierarchies which we treated in the first two sections of this note. Therefore, we already have tentative parameters for the half-Woodrow structure, namely:

$$\alpha\ = .71 \qquad \beta_1 = .66$$
$$\alpha_5 = .76 \qquad \beta_2 = .58$$
$$\alpha_6 = .83 \qquad \beta_3 = .41$$
$$\alpha_7 = .84 \qquad \beta_4 = .32$$
$$\alpha_8 = .87 \qquad \beta\ = .00$$

where α represents the first four loadings (all equal) on the first factor and β represents the last four loadings (again, all equal) on the second factor.

These loadings, of course, were derived exclusively from the two apical triangles. To this point we have taken no account of the correlations in the central block. To do so, however, we need only generate the block correlations from our preliminary loadings and see what sort of a fit we get. If the fit is good, our preliminary loadings become final. If it isn't, we must adjust the factor loadings so as to compromise the errors which they generate in the apical traingles and in the central block. In the case at hand the fit is not good. But the remedy is easy. All we need do is assign β a somewhat higher value than zero. If we let β equal .20 and keep all the other loadings just what they were, we get the theoretical correlations which appear in Table A6, and a good fit. Of course, it might not have been possible to fit the central block and the apical triangles with a single set of loadings. If this had been the case, we would have had to reject the half-Woodrow hypothesis.

TABLE A6

SOME THEORETICAL CORRELATIONS GENERATED BY A HALF-WOODROW TWO-FACTOR STRUCTURE

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.88	.77	.71	.67	.72	.73	.75
X_2			.74	.69	.66	.71	.71	.73
X_3				.63	.62	.67	.68	.70
X_4					.60	.65	.66	.68
X_5						.67	.68	.70
X_6							.74	.76
X_7								.77
X_8								

FULL–WOODROW TWO–FACTOR STRUCTURE

Fitting a full-Woodrow structure depends upon one fact: If a correlation matrix is derived from a full-Woodrow structure, it is always possible to transform the correlations into a three-cornered pattern and to fit the transformed correlations with a half-Woodrow structure. Suppose, for example, that we multiply the loadings a_i and b_i in a full-Woodrow structure by k_i, where:

$$k_i = \frac{a_i}{a_i}$$

for the first $n/2$ variables and

$$k_2 = \frac{b_n}{b_i}$$

for the last $n/2$ variables. The result will be a half-Woodrow structure, with the loadings (in the eight-variable case):

$$\begin{array}{llll}
\alpha_1 = a_1 & \alpha_5 = a_5\left(\dfrac{b_8}{b_5}\right) & \beta_1 = b_1 & \beta_5 = b_8 \\
\alpha_2 = a_1 & \alpha_6 = a_6\left(\dfrac{b_8}{b_6}\right) & \beta_2 = b_2\left(\dfrac{a_1}{a_2}\right) & \beta_6 = b_8 \\
\alpha_3 = a_1 & \alpha_7 = a_7\left(\dfrac{b_8}{b_7}\right) & \beta_3 = b_3\left(\dfrac{a_1}{a_3}\right) & \beta_7 = b_8 \\
\alpha_4 = a_1 & \alpha_8 = a_8 & \beta_4 = b_4\left(\dfrac{a_1}{a_4}\right) & \beta_8 = b_8
\end{array}$$

And the correlations which this new structure generates, call them r_{ij}', will bear a simple relationship to the correlations of the original matrix, namely:

$$r_{ij}' = r_{ij}k_ik_j$$

It follows that to fit a full-Woodrow structure we must do three things: transform the superdiagonal form into a three-cornered pattern, fit the three-cornered pattern with a half-Woodrow, and then transform the loadings for the half-Woodrow back to the original full-Woodrow structure.

Our first task, then, is to find values for the k_i which are inversely proportional to the full-Woodrow loadings, a_1, a_2, a_3, a_4, and b_5, b_6, b_7, b_8. The word "proportional" is important, because it means that we have a great deal of latitude in transforming the superdiagonal form; more specifically, it means that the transformed correlations may be big, small, or middling, without any loss of rigor. As long as the new correlations can be fit by a half-Woodrow structure, whatever their absolute magnitude may be, then we can fit the original correlations with a full-Woodrow structure. But even so, how do we find a set of k_i which will do the trick?

Consider the correlations in the central block of a three-cornered pattern. They take this form:

$$\begin{array}{lllll}
 & X_5 & X_6 & X_7 & X_8 \\
X_1 & (\alpha\alpha_5+\beta\beta_1) & (\alpha\alpha_6+\beta\beta_1) & (\alpha\alpha_7+\beta\beta_1) & (\alpha\alpha_8+\beta\beta_1) \\
X_2 & (\alpha\alpha_5+\beta\beta_2) & (\alpha\alpha_6+\beta\beta_2) & (\alpha\alpha_7+\beta\beta_2) & (\alpha\alpha_8+\beta\beta_2) \\
X_3 & (\alpha\alpha_5+\beta\beta_3) & (\alpha\alpha_6+\beta\beta_3) & (\alpha\alpha_7+\beta\beta_3) & (\alpha\alpha_8+\beta\beta_3) \\
X_4 & (\alpha\alpha_5+\beta\beta_4) & (\alpha\alpha_6+\beta\beta_4) & (\alpha\alpha_7+\beta\beta_4) & (\alpha\alpha_8+\beta\beta_4)
\end{array}$$

which the reader will immediately recognize as a perfectly additive linear pattern, as these terms are used in analysis of variance. What we seek, therefore, are values for the k_i which will transform the central block of the superdiagonal form into this pattern. Now, we know from the nature of the Woodrow structure that:

$$\alpha_5 \leq \alpha_6 \leq \alpha_7 \leq \alpha_8$$

and

$$\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$$

which means that the biggest r_{ij}' in the central block will be in the upper right-hand corner. The correlations will fall off as we go either to the left or down, until we come to the smallest r_{ij}' in the central block, which will be in the lower left-hand corner. We also know, however, that the correlations in the central block of a superdiagonal form have just the opposite pattern: The biggest correlation is in the lower left-hand corner and the smallest in the upper right-hand corner. It follows that the multiplicative constants, k_i, must obey these inequalities:

$$k_1 \geq k_2 \geq k_3 \geq k_4$$

and

$$k_5 \leq k_6 \leq k_7 \leq k_8$$

where the first four constants are multiplied only into the rows of the central block and the last four only into the columns. Knowing this much there is no great difficulty in transforming the correlations so that they have the prescribed pattern.

Take the correlations in Table A5, for example. By now it is quite obvious that the correlations in Table A5 were derived from the original correlations for the Two-Hand Coordination Test. To produce the correlations of Table A5 each correlation r_{ij} in Table 1 was multiplied by k_i and again by k_j where:

$$\begin{array}{ll} k_1 = 1.17 & k_5 = .81 \\ k_2 = .95 & k_6 = .86 \\ k_3 = .86 & k_7 = .88 \\ k_4 = .81 & k_8 = .90 \end{array}$$

All we have to do now, therefore, is to divide the loadings of the half-Woodrow structure, the α_i and β_i, by their corresponding k_{ij} and we have our tentative parameters for the original full-Woodrow structure, viz.:

$$\begin{array}{llll} a_1 = .61 & a_5 = .94 & b_1 = .56 & b_5 = .25 \\ a_2 = .75 & a_6 = .97 & b_2 = .61 & b_6 = .23 \\ a_3 = .83 & a_7 = .95 & b_3 = .48 & b_7 = .23 \\ a_4 = .88 & a_8 = .97 & b_4 = .40 & b_8 = .22 \end{array}$$

These values do not conform entirely to the requirements of the full-Woodrow structure. b_1 should not be less than b_2; nor should a_7 be less than a_6. Also, the fit, though good, can be improved. However, we can make adjustments for these things; and when we do, we get the loadings which are cited in the text and the correlations which are reproduced in Table 2.

DOUBLE ACCUMULATION STRUCTURE

If a correlation matrix is generated by a double accumulation structure, then:

$$r_{ij} \approx r_{1n} + (r_{ij} - r_{1n}) + (r_{in} - r_{1n})$$

If the fit is perfect, the relationship is exact. In other words, the correlations generated by a double accumulation structure have the same additive linear pattern which we remarked in the central block of the half-Woodrow pattern. This circumstance makes it possible for us to fit the correlations under a double accumulation hypothesis without ever specifying the parameters of the structure, i.e., the variances of the factors.

Let's take a few correlations from the matrix in Table 1 and see how well they satisfy this linear, additive relationship. We have:

$$\begin{aligned} r_{23} &= .87 \approx .70 + (.07 + .12) = .89 \\ r_{34} &= .91 \approx .70 + (.04 + .06) = .90 \\ r_{45} &= .91 \approx .70 + (.03 + .18) \\ r_{37} &= .85 \approx .70 + (.01 + .16) = .87 \end{aligned}$$

It's quite obvious from these relationships that Table 1 can be accounted for as a double accumulation structure. All we have to do is to treat the correlations in the first row and last column as theoretical correlations and adjust them until we get a good fit, which, in the present case, is hardly much of a chore. Once they are made, we have the theoretical correlations, ρ_{ij}, which appear in Table 3, of the text.

It remains, of course, to determine the variances of the factors. In a double accumulation structure the first and last variables in the matrix have only one factor in common and that one factor is common to all of the variables in the matrix. It follows that the variance of this one factor equals r_{1n}. Now, apart from this general factor, a double accumulation structure contains $2(n - 2)$ other factors, which are divided into two accumulation structures, each one containing $(n - 2)$ factors. It happens, however, that the theoretical correlations in the first row of a double accumulation pattern are determined by one of these structures only, and the theoretical correlations in the last column are determined by the other, exclusively. This means that we can reel off the factor variances in a very simple way, once we know the theoretical correlations. If the factors are numbered the way they are in

FULL–WOODROW TWO–FACTOR STRUCTURE

Fitting a full-Woodrow structure depends upon one fact: If a correlation matrix is derived from a full-Woodrow structure, it is always possible to transform the correlations into a three-cornered pattern and to fit the transformed correlations with a half-Woodrow structure. Suppose, for example, that we multiply the loadings a_i and b_i in a full-Woodrow structure by k_i, where:

$$k_i = \frac{a_i}{a_i}$$

for the first $n/2$ variables and

$$k_2 = \frac{b_n}{b_i}$$

for the last $n/2$ variables. The result will be a half-Woodrow structure, with the loadings (in the eight-variable case):

$$\begin{array}{llll}
\alpha_1 = a_1 & \alpha_5 = a_5\left(\dfrac{b_8}{b_5}\right) & \beta_1 = b_1 & \beta_5 = b_8 \\
\alpha_2 = a_1 & \alpha_6 = a_6\left(\dfrac{b_8}{b_6}\right) & \beta_2 = b_2\left(\dfrac{a_1}{a_2}\right) & \beta_6 = b_8 \\
\alpha_3 = a_1 & \alpha_7 = a_7\left(\dfrac{b_8}{b_7}\right) & \beta_3 = b_3\left(\dfrac{a_1}{a_3}\right) & \beta_7 = b_8 \\
\alpha_4 = a_1 & \alpha_8 = a_8 & \beta_4 = b_4\left(\dfrac{a_1}{a_4}\right) & \beta_8 = b_8
\end{array}$$

And the correlations which this new structure generates, call them r_{ij}', will bear a simple relationship to the correlations of the original matrix, namely:

$$r_{ij}' = r_{ij}k_ik_j$$

It follows that to fit a full-Woodrow structure we must do three things: transform the superdiagonal form into a three-cornered pattern, fit the three-cornered pattern with a half-Woodrow, and then transform the loadings for the half-Woodrow back to the original full-Woodrow structure.

Our first task, then, is to find values for the k_i which are inversely proportional to the full-Woodrow loadings, a_1, a_2, a_3, a_4, and b_5, b_6, b_7, b_8. The word "proportional" is important, because it means that we have a great deal of latitude in transforming the superdiagonal form; more specifically, it means that the transformed correlations may be big, small, or middling, without any loss of rigor. As long as the new correlations can be fit by a half-Woodrow structure, whatever their absolute magnitude may be, then we can fit the original correlations with a full-Woodrow structure. But even so, how do we find a set of k_i which will do the trick?

Consider the correlations in the central block of a three-cornered pattern. They take this form:

$$\begin{array}{lllll}
 & X_5 & X_6 & X_7 & X_8 \\
X_1 & (\alpha\alpha_5+\beta\beta_1) & (\alpha\alpha_6+\beta\beta_1) & (\alpha\alpha_7+\beta\beta_1) & (\alpha\alpha_8+\beta\beta_1) \\
X_2 & (\alpha\alpha_5+\beta\beta_2) & (\alpha\alpha_6+\beta\beta_2) & (\alpha\alpha_7+\beta\beta_2) & (\alpha\alpha_8+\beta\beta_2) \\
X_3 & (\alpha\alpha_5+\beta\beta_3) & (\alpha\alpha_6+\beta\beta_3) & (\alpha\alpha_7+\beta\beta_3) & (\alpha\alpha_8+\beta\beta_3) \\
X_4 & (\alpha\alpha_5+\beta\beta_4) & (\alpha\alpha_6+\beta\beta_4) & (\alpha\alpha_7+\beta\beta_4) & (\alpha\alpha_8+\beta\beta_4)
\end{array}$$

which the reader will immediately recognize as a perfectly additive linear pattern, as these terms are used in analysis of variance. What we seek, therefore, are values for the k_i which will transform the central block of the superdiagonal form into this pattern. Now, we know from the nature of the Woodrow structure that:

$$\alpha_5 \leq \alpha_6 \leq \alpha_7 \leq \alpha_8$$

and

$$\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$$

which means that the biggest r_{ij}' in the central block will be in the upper right-hand corner. The correlations will fall off as we go either to the left or down, until we come to the smallest r_{ij}' in the central block, which will be in the lower left-hand corner. We also know, however, that the correlations in the central block of a superdiagonal form have just the opposite pattern: The biggest correlation is in the lower left-hand corner and the smallest in the upper right-hand corner. It follows that the multiplicative constants, k_i, must obey these inequalities:

$$k_1 \geq k_2 \geq k_3 \geq k_4$$

and

$$k_5 \leq k_6 \leq k_7 \leq k_8$$

where the first four constants are multiplied only into the rows of the central block and the last four only into the columns. Knowing this much there is no great difficulty in transforming the correlations so that they have the prescribed pattern.

Take the correlations in Table A5, for example. By now it is quite obvious that the correlations in Table A5 were derived from the original correlations for the Two-Hand Coordination Test. To produce the correlations of Table A5 each correlation r_{ij} in Table 1 was multiplied by k_i and again by k_j where:

$k_1 = 1.17$ $k_5 = .81$
$k_2 = .95$ $k_6 = .86$
$k_3 = .86$ $k_7 = .88$
$k_4 = .81$ $k_8 = .90$

All we have to do now, therefore, is to divide the loadings of the half-Woodrow structure, the α_i and β_i, by their corresponding k_{ij} and we have our tentative parameters for the original full-Woodrow structure, viz.:

$a_1 = .61$ $a_5 = .94$ $b_1 = .56$ $b_5 = .25$
$a_2 = .75$ $a_6 = .97$ $b_2 = .61$ $b_6 = .23$
$a_3 = .83$ $a_7 = .95$ $b_3 = .48$ $b_7 = .23$
$a_4 = .88$ $a_8 = .97$ $b_4 = .40$ $b_8 = .22$

These values do not conform entirely to the requirements of the full-Woodrow structure. b_1 should not be less than b_2; nor should a_7 be less than a_6. Also, the fit, though good, can be improved. However, we can make adjustments for these things; and when we do, we get the loadings which are cited in the text and the correlations which are reproduced in Table 2.

DOUBLE ACCUMULATION STRUCTURE

If a correlation matrix is generated by a double accumulation structure, then:

$$r_{ij} \approx r_{1n} + (r_{ij} - r_{1n}) + (r_{in} - r_{1n})$$

If the fit is perfect, the relationship is exact. In other words, the correlations generated by a double accumulation structure have the same additive linear pattern which we remarked in the central block of the half-Woodrow pattern. This circumstance makes it possible for us to fit the correlations under a double accumulation hypothesis without ever specifying the parameters of the structure, i.e., the variances of the factors.

Let's take a few correlations from the matrix in Table 1 and see how well they satisfy this linear, additive relationship. We have:

$$r_{23} = .87 \approx .70 + (.07 + .12) = .89$$
$$r_{34} = .91 \approx .70 + (.04 + .06) = .90$$
$$r_{45} = .91 \approx .70 + (.03 + .18)$$
$$r_{37} = .85 \approx .70 + (.01 + .16) = .87$$

It's quite obvious from these relationships that Table 1 can be accounted for as a double accumulation structure. All we have to do is to treat the correlations in the first row and last column as theoretical correlations and adjust them until we get a good fit, which, in the present case, is hardly much of a chore. Once they are made, we have the theoretical correlations, ρ_{ij}, which appear in Table 3, of the text.

It remains, of course, to determine the variances of the factors. In a double accumulation structure the first and last variables in the matrix have only one factor in common and that one factor is common to all of the variables in the matrix. It follows that the variance of this one factor equals r_{1n}. Now, apart from this general factor, a double accumulation structure contains $2(n - 2)$ other factors, which are divided into two accumulation structures, each one containing $(n - 2)$ factors. It happens, however, that the theoretical correlations in the first row of a double accumulation pattern are determined by one of these structures only, and the theoretical correlations in the last column are determined by the other, exclusively. This means that we can reel off the factor variances in a very simple way, once we know the theoretical correlations. If the factors are numbered the way they are in the text, we get:

$$\sigma_{f_2}^2 = \rho_{17} - \rho_{18}$$
$$\sigma_{f_3}^2 = \rho_{16} - \rho_{17}$$
$$\cdots\cdots\cdots\cdots$$
$$\sigma_{f_7}^2 = \rho_{12} - \rho_{13}$$

and

$$\sigma_{f_{13}}^2 = \rho_{28} - \rho_{18}$$
$$\sigma_{f_{12}}^2 = \rho_{38} - \rho_{28}$$
$$\cdots\cdots\cdots\cdots$$
$$\sigma_{f_8}^2 = \rho_{78} - \rho_{68}$$

From these formulae I derived the values of the $\sigma_{f_i}^2$ which appear in the text.

SIMPLEX

In a Woodrow structure the parameters are all of the same sort, namely factor loadings; similarly, in the double accumulation structure, the parameters are also of the same sort, in this case factor variances. But in the simplex there are two different kinds of structural parameters. Under a simplicial hypothesis, the theoretical correlations:

$$\rho_{ij} = \frac{c_j}{c_i}\sqrt{h_i^2 h_j^2} \qquad (i < j)$$

where the c_i are the complexities and the h_i^2 are the communalities. This parametric duality creates a corresponding duality in the fitting procedures: We can specify the communalities first or we can specify the complexities first. The older method is to specify the communalities first. So I will begin there.

Communality Method

Using the communality method we fit a simplex in the following manner. First, we specify tentative values for the communalities. Then, the correlations are corrected for attenuation (by the communalities) and the resulting, unattenuated correlations fit with a perfect simplicial form.[A1] If the unattenuated correlations won't submit to a simplicial fit, it may be that the variables which enter into the matrix are not structured simplicially. It may also be that we used the wrong communalities. So we adjust the communalities. If, by these adjustments, we can bring the unattenuated correlations into simplicial form or an approximation to it, we have our fit. If we can't, then the simplicial hypothesis must be rejected.

It's quite obvious from this account that a great deal hinges on the original choice of communalities. If the matrix has simplicial form, then sooner or later we will find a fit. But if the values we posit for the communalities are sufficiently far off, it may be much later. It would be nice if we had some way of knowing at the beginning what these communalities would turn out to be. Unfortunately, in advance of trying them out, the best we can do is to specify limits for the communalities. It would not be reasonable, for example, to set the communality of a variable higher than its known reliability. Nor would it be resonable to set it so low that the correlation between two variables, after correcting for the communalities, became greater than unity. But within these limits there is no way of knowing in advance what the best-fitting values for the communalities are.

Nevertheless, we can still posit values for the communalities and try them out. And this is just what we do. In practice, there are three basic ways of proceeding. One, we can set the communality of the *i*th variable equal to the highest correlation it has in the matrix. Two, we can equate the communalities with one kind or another of reliability (if we have any information on this count). Or three, we can set the communalities all equal to a given value, .90, say, or .80. No one of these procedures has any a priori claim to being right. But they do represent so many ways of approaching the problem, and in my experience they are good ways. I have fit a great many simplexes, and one or another of these approaches, with minor adjustments, has always led to a good fit, if one existed.

In fitting the matrix for the Two-Hand Coordination Test I set the communali-

---

[A1] The term "perfect simplicial form" is defined on the next page. For a fuller discussion of the simplex and its terminology, see Jones (1959, pp. 11–15).

ties of the trials equal to the largest correlation they had in the matrix. The next step was to correct every correlation in the matrix for its attenuation by the remaining, specific variance. In other words, I divided r_{ij} by $\sqrt{h_{ij}h_j^2}$. The matrix of unattenuated coefficients which resulted appears in Table A7.

Now, once the correlations have been corrected for attenuation the equations for the simplex become:

$$\begin{aligned} x_1 &= f_1 + f_2 + \cdots + f_{n-1} + f_n \\ x_2 &= \phantom{f_1 + {}} f_2 + \cdots + f_{n-1} + f_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{n-1} &= \phantom{f_1 + f_2 + \cdots + {}} f_{n-1} + f_n \\ x_n &= \phantom{f_1 + f_2 + \cdots + f_{n-1} + {}} f_n \end{aligned}$$

or just what they were originally, except that the specific factors are gone. In this form, the simplex lays down very demanding requirements on the correlations. If the unattenuated correlations r_{ij}', really do form a simplex, then they must obey the fundamental law of "perfect simplicial form," i.e.:

$$r_{ik}' = r_{ij}'r_{jk}'$$

For example, if $r_{12}' = .90$ and $r_{23}' = .80$, then r_{13}' must equal .72. Clearly, if this is the case, the matrix will have the superdiagonal form—and then some.

Of course, in practice the unattenuated coefficients never do have simplicial form, not exactly. Take the correlations of Table A7, for example. r_{12}' times r_{23}' equals .93, which almost equals $r_{13}' =$ .91, but not exactly. Our task, therefore, is to find a theoretical simplicial form which comes as close as possible to the unattenuated pattern. In the case of Table A7 this task is easy. The correlations, which appear in Table A8 have perfect simplicial form, and, as the reader can see, they constitute a very close approximation to the unattenuated empirical correlations.

TABLE A7

UNATTENUATED EMPIRICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COMMUNALITY METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.95	.91	.87	.86	.83	.83	.81
X_2			.98	.98	.94	.91	.91	.91
X_3				1.00	.98	.95	.92	.93
X_4					1.00	.96	.93	.96
X_5						.97	.98	.98
X_6							1.00	1.00
X_7								1.00
X_8								

TABLE A8

UNATTENUATED THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COMMUNALITY METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.93	.90	.88	.86	.83	.83	.83
X_2			.97	.95	.93	.90	.90	.90
X_3				.98	.96	.93	.93	.93
X_4					.98	.95	.95	.95
X_5						.97	.97	.97
X_6							1.00	1.00
X_7								1.00
X_8								

To obtain the theoretical correlations, c_{ij}, which I presented in Table 4 of the text I simply multiplied the correlations in Table A8 by the same communalities which I used to unattenuate the empirical correlations.

But what about the complexities? Here we are, with the ρ_{ij} already derived, and we still haven't said anything about the complexities. The answer, of course, is that, without realizing it, perhaps, we have already posited the complexities. The complexity of the *i*th trial is nothing other than its theoretical correlation, *before attenuation*, with the first trial in the series; in other words, $c_j = c_{ij}$. In the case at hand, therefore, the complexities are the correlations in the first row of Table A8.[A2]

[A2] The virtue of the complexities is that, if we know them, then by the fundamental law of perfect simplicial form we know all of the c_{ij} and can, therefore, if we also know the communalities, generate all of the theoretical correlations ρ_{ij}. However, the complexities are not alone in possessing this property. The gradations, i.e., the correlations $c_{i,i+1}$, are also capable of generating the entire set

Covariance Method[A3]

The full equations for the simplex (with progressive simplification) are:

$$x_i = k_i(f_i + \cdots + f_n + s_i)$$

where the k_i are arbitrary constants. In the usual presentation of the simplex, as in the text of this paper, these constants are dropped, since they are of no consequence as long as we have to do with correlations only. The moment, however, that we concern ourselves with variances or with covariances we must dredge these little things up and be most meticulous about them.

Under a simplicial hypothesis, the covariances between the variables:

$$\text{cov}_{ij} = k_i k_j(\sigma_{f_j}{}^2 + \cdots + \sigma_{f_n}{}^2)$$

Suppose, now, that we multiply all of the entries in the ith row by $\frac{k_o}{k_i}$, where k_o is another arbitrary constant. The new entries:

$$\text{cov}_{ij}{}^1 = k_o k_j(\sigma_{f_j}{}^2 \cdots + \sigma_{f_n}{}^2$$

And if we multiply the columns by the same quantities, we get the entries:

$$\text{cov}_{ij}{}^{11} = k_o{}^2(\sigma_{f_j}{}^2 + \cdots + \sigma_{f_n}{}^2$$

From here we get the unattenuated, theoretical correlations c_{ij} immediately, because:

$$c_{ij}{}^2 = \frac{(\sigma_{f_j}{}^2 + \cdots + \sigma_{f_n}{}^2)}{(\sigma_{f_i}{}^2 + \cdots + \sigma_{f_n}{}^2)}$$

The only problem is to find the $\frac{k_o}{k_i}$.

TABLE A9

THE COVARIANCE MATRIX FOR THE TWO-HAND COORDINATION TEST

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		323	331	329	376	353	360	375
X_2			477	484	546	510	523	528
X_3				536	617	575	580	584
X_4					640	593	598	613
X_5						702	726	729
X_6							718	722
X_7								751
X_8								

This problem, happily, is easily solved. Once we have multiplied through the rows by the $\frac{k_o}{k_i}$, then all of the entries in any one column are equal. So suppose we set all of the entries in the last column equal to some arbitrary amount A, by multiplying the rows by:

$$\frac{A}{\text{cov}_{in}} = \frac{k_o}{k_i}\left(k_o = \frac{A}{k_n \sigma_{f_n}{}^2}\right)$$

If the simplicial hypothesis is correct, then all of the entries in any one of the other columns should be equal too, or close to it. If this is not the case, we can reject the simplicial hypothesis. If it is, however, we have the $\frac{k_o}{k_i}$ we sought.

In Table A9 the covariance matrix for the Two-Hand Coordination Test is presented. Setting all the entries in the last column equal to 300 (except for rounding errors), we get these values for our $\frac{k_o}{k_i}$.

$$\frac{k_o}{k_1} = .84 \quad \frac{k_o}{k_3} = .51 \quad \frac{k_o}{k_5} = .41 \quad \frac{k_o}{k_7} = .40$$

$$\frac{k_o}{k_o} = .57 \quad \frac{k_o}{k_4} = .49 \quad \frac{k_o}{k_6} = .42$$

Note that we get no value for $\frac{k_o}{k_8}$. When the rows are multiplied by these values, we get the matrix which appears in Table A10. And since the columns conform approximately to the simplicial requirements, we know that we are on the right

---

of c_{ij} and so are several other subsets of the c_{ij}. For a discussion of this point, see Jones (1959, p. 12).

[A3] The first person to mention to me the possibilities of the covariance matrix in fitting the simplex to a correlation matrix was Ardie Lubin. My original reaction was very negative and in the "Simplex Theory" (Jones, 1959, pp. 79–80) I had some rather hard things to say about his suggestion. I have no course now, therefore, but to apologize to him for my intemperance: he was right and I was wrong.

TABLE A10

THE COVARIANCE MATRIX FOR THE TWO-HAND COORDINATION TEST AFTER MULTIPLYING THROUGH THE ROWS BY $\frac{k_o}{k_i}$

Trial	X_1	X_2	X_4	X_4	X_5	X_6	X_7	X_8
X_1		271	278	276	316	297	302	300
X_2			272	275	311	291	298	301
X_3				273	315	293	296	298
X_4					314	291	293	300
X_5						288	298	299
X_6							302	303
X_7								300
X_8								

track. Now, multiplying through the columns by these same values we get the matrix which appears in Table A11. And from this table, taking the average down the columns, we get:

$$k_o^2(\sigma_{f_2}^2 + \cdots + \sigma_{f_8}^2) = 154$$
$$k_o^2(\sigma_{f_3}^2 + \cdots + \sigma_{f_8}^2) = 141$$
$$k_o^2(\sigma_{f_4}^2 + \cdots + \sigma_{f_8}^2) = 135$$

$$k_o^2(\sigma_{f_5}^2 + \cdots + \sigma_{f_8}^2) = 129$$
$$k_o^2(\sigma_{f_6}^2 + \sigma_{f_7}^2 + \sigma_{f_8}^2) = 123$$
$$k_o^2(\sigma_{f_7}^2 + \sigma_{f_8}^2) = 119$$

From these figures we immediately obtain the matrix of theoretical, unattenuated correlation coefficients (minus the first row and the last column) which appears in Table A12. It would have made no difference, by the way, if we had equated the rows first and then multiplied through the columns. We would have ended up with the same values for the c_{ij}.

TABLE A11

THE COVARIANCE MATRIX FOR THE TWO-HAND COORDINATION TEST AFTER MULTIPLYING THROUGH THE ROWS AND THE COLUMNS BY $\frac{k_o}{k_i}$

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		154	142	135	130	125	121	
X_2			139	135	128	122	119	
X_3				134	129	123	118	
X_4					129	122	117	
X_5						121	119	
X_6							121	
X_7								
X_8								

Let's not worry for the time being about the first row and the last column and proceed directly to the communalities. If a simplicial fit exists, we know that:

$$r_{ij} \approx c_{ij}\sqrt{h_i^2 h_j^2}$$

If, therefore, we divide the r_{ij} by our already obtained c_{ij}, the resulting matrix should be a Spearman unit hierarchy, where the loadings for the several variables equal $\sqrt{h_i^2}$. Carrying out this process we find that:

$$h_2^2 = .92 \qquad h_4^2 = .92 \qquad h_6^2 = .94$$
$$h_3^2 = .92 \qquad h_5^2 = .94 \qquad h_7^2 = .94$$

TABLE A12

UNATTENUATED THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COVARIANCE METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1								
X_2			.96	.94	.92	.89	.88	
X_3				.98	.96	.93	.92	
X_4					.98	.95	.94	
X_5						.97	.96	
X_6							.98	
X_7								
X_8								

It only remains to fill things in for the first and last variables.

The correlations in the first row:

$$r_{1j} \approx c_{12}c_{2j}\sqrt{h_1^2 h_j^2}$$

or

$$r_{1j} \approx l_1 c_{2j}\sqrt{h_j^2}$$

if we set

$$l_1 = c_{12}\sqrt{h_1^2}$$

In other words, given what we already know, i.e., c_{2j} and $\sqrt{h_j^2}$, for $2 \leq j \leq n - 1$, the empirical correlations r_{1j} are generated by a *single* parameter. In the same way, the empirical correlations in the last column:

$$r_{in} \approx l_2 c_{i1(n-1)} \sqrt{h_i^2}$$

where

$$l_n = c_{(n-1),n} \sqrt{h_n^2}$$

and finally

$$r_{1n} = l_1 l_n c_{27}$$

All we have to do, therefore, is to obtain tentative values for l_1 and l_n. The most natural way to do this is to set:

$$l_1 = \frac{r_{12}}{\sqrt{h_2^2}},$$

$$l_n = \frac{r_{(n-1)n}}{\sqrt{h_{(n-1)}^2}}$$

and then solve for l_1 and l_n. In the present case, we get:

$$l_1 = .82,$$
$$l_8 = .97$$

Altogether, then, we can pick any values we like for c_{12} and h_1^2, provided only that:

$$c_{12} \sqrt{h_1^2} = .82$$

And in the same way we can pick any values we like for c_{78} and h_8^2, provided only that:

$$c_{78} \sqrt{h_8^2} = .97$$

Suppose, therefore, that we set:

$$c_{12} = .90 \qquad c_{78} = .99$$
$$h_1^2 = .81 \qquad h_8^2 = .96$$

Remembering that the complexities are simply the c_{ij}, we now have both our communalities and our complexities, and we can generate our theoretical correlations. The ρ_{ij} for the Two-Hand Coordination Test appear in Table A13, and, as the reader can see, we have a very good fit. Also, if the reader will check the parameters obtained by this method with those obtained by the communality method, he will see that they are essentially the same.

TABLE A13

ATTENUATED THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COVARIANCE METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.79	.75	.73	.72	.70	.69	.69
X_2			.88	.86	.86	.83	.82	.82
X_3				.90	.89	.86	.86	.86
X_4					.91	.88	.87	.87
X_5						.91	.90	.90
X_6							.92	.92
X_7								.94
X_8								

In conclusion, let me say a word or two about the comparative merits of the communality and the covariance methods. Ordinarily, the communality method is quicker and involves many fewer computations. But it can go awry. If the communalities of the variables happen not to coincide, even approximately, with the tentative values which are given us by one or another of our three methods of approach, we can waste an awful lot of time looking around for a fit. The covariance method, in comparison, is sure. If a fit exists. you can always get it with the covariance method.

Psychological Review
1962, Vol. 69, No. 4, 295–305

NORMATIVE AND IPSATIVE MEASUREMENT IN PSYCHOLOGY[1]

DONALD M. BROVERMAN

Worcester State Hospital, Massachusetts

In an earlier paper (Broverman, 1961) certain differences between factors obtained from standard R factor analysis methods (correlating tests) and Q factor analysis techniques (correlating people) were illustrated and attributed to the different types of scores used by the two techniques. The usual R analysis uses "normative" scores, or scores which reflect deviations of individual performances from the *group's* mean level of performance; while Q factor analysis typically uses "ipsative" scores, or scores which reflect deviations of particular performances from each *individual's* mean level of performance. Normative scores, then, are distributed about a norm characterizing a group, while ipsative scores are based on norms of individuals.

The present paper will attempt to show that normative and ipsative possibilities of measurement occur not only in factor analysis, but, also, in most efforts to quantify behavior, whether in purely theoretical terms or in actual experimental situations. Accordingly, the assumptions involved in each mode of measurement, and examples of these assumptions in current psychological theories will be pointed out. The paper will then examine the manner in which standard statistical procedures cope, or fail to cope, with these different aspects of behavior. With respect to this last point, the classical correlational procedures will be shown to confound normative and ipsative aspects of behavior. Hence, the final portion of this paper will present correlational techniques for dealing with these distinctions.

ASSUMPTIONS OF THE NORMATIVE MEASUREMENT ORIENTIATION

The normative measurement orientation assumes that behavioral variations between individuals may be ordered on parameters which exist within the species' universe of variation. Thus, normative measurement procedures distribute scores about the norm of a group, sample, or population. In normative measurement, then, the behaviors of individuals are ordered on dimensions which transcend the individual, i.e., the behavior of a given individual is but a point or mark on a normative scale. This approach assumes that certain individual differences in behavior are due to an entity or trait which all members of the population possess in varying amounts. For example, Spearman (1927) assumes that individual differences in intelligence reflect the different amounts of mental energy possessed by different individuals.

It should be noted that intraindividual variation in parts of normative traits are not considered in normative measurement. Accordingly, less than perfect correlations between several manifestations of the same normative trait are accounted for by assuming (*a*) that the surface behaviors showing the trait are phenotypically complex manifestations of more than one underlying genotypical, normative trait, and/or (*b*) that part of the discrepancy is due to measurement error. Thus, normative dimensions are treated as unitary, internally undifferentiated variables whose variation occurs across individuals.

---

[1] Preparation of this paper was facilitated by the Dementia Praecox Research Project, Worcester State Hospital, and research grant (M-896), from the National Institute of Mental Health, United States Public Health Service. The writer prepared this paper while a United States Public Health Postdoctoral Fellow in Clinical Psychology.

ASSUMPTIONS OF THE IPSATIVE MEASUREMENT ORIENTATION

The ipsative measurement orientation, on the other hand, assumes that each individual is, himself, a universe within which variations in behavior occur. Accordingly, the different behaviors of an individual are ordered on dimensions or scales within each individual, e.g., the habit strengths of different responses in an individual may be scaled and compared. Particular behaviors of the individual, then, are assigned marks or scores on ipsative scales.

The ipsative measurement approach frequently assumes that relationships between psychic entities in an individual determine his behavior. For instance, Spence (1954) argues that when incompatible response tendencies are simultaneously aroused in an individual, the response possessing the strongest habit strength will be the one the individual will make.

It should be noted that often the ipsative orientation is not concerned with the unique aspects of single individuals as in the idiographic method advocated by Allport (1937). On the contrary, ipsative relationships are frequently ascribed in theory to an entire population, i.e., the rules or laws between internal psychic events may be generalized to all humans. Thus, it is not contradictory to seek commonly occurring ipsative relationships. Indeed, this is the goal of many personality theories.

The ipsative measurement orientation, then, is concerned with quantitative relationships between psychic functions within individuals. It frequently assumes that such intrapsychic relationships play an important role in the determination of behavior.

IPSATIVE MEASUREMENT PROCEDURES

A property of all scales, both ipsative and normative, is that the scale scores must be in the same units. Hence, in ipsative measurement, the various behaviors of a given individual must be expressed in terms of some common unit if scale comparisons are to be made. Numerous examples in psychological theory illustrating this procedure are available. For instance, Spearman (1927), who maintains that intelligence has certain subfunctions such as eduction of relationships, reproduction, etc., posits that a large expenditure of mental energy in one subfunction leaves less energy available for other subfunctions. In this example, the common unit of the different subfunctions is mental energy. The energy levels of the two subfunctions, therefore, may be scaled and compared. Similarly Freud (1935) posits conflict between different psychic institutions, e.g., id versus ego, the outcome of which depends upon their respective strengths. Once more the common denominator is some sort of mental energy.

Thus, in ipsative scaling procedures, different behaviors are reduced to, or expressed in terms of some common unit, and vary around a norm established within the individual. The norm, of course, is also expressed in these same units.

Different ways of establishing ipsative

norms within individuals exist. One method is by working with single, isolated individuals. For example, in Stephenson's (1953) *Q* sort each individual produces his own distribution of responses to a set of stimuli. The typical unit of variation in such *Q* sorts is the degree of "likeness to me." Another example would be measures of physiological reactivity under different conditions which are compared to a base-rate of that activity. In these examples, the norms are derived solely from a given individual's behaviors. What other individuals do has no effect upon the scales or norms so established. Hence, this may be called a "pure" ipsative measurement method.

A second way of establishing ipsative norms within individuals starts with normative considerations. For instance, the performances of a number of subjects might be normatively ranked for each test in a test battery. A given subject would then have a distribution of normative ranks—one rank for each of the tests in the battery. The subject's general level of performance for the battery, i.e., the subject's average normative rank standing, would then constitute the subject's ipsative norm and deviations of the subject's normative rank standings on particular tests from this ipsative norm would represent ipsative scores. The ipsative scores, in this case, are in units of deviations of group standings on various tasks from the subject's average group standing. Ipsative scores derived in this manner may be termed "double-centered" scores, i.e., they are centered or distributed first about a group norm, then about a norm of the individual.

Thus, unlike either the pure ipsative method or the normative approach, the double-centered ipsative approach takes into consideration both normative and ipsative aspects of variation.

MODES OF MEASUREMENT VERSUS PARAMETERS OF BEHAVIOR

Measurement operations refer to certain ways of assigning values to events or data on hypothetical parameters. Measurement-wise, however, one may develop several different techniques for establishing such values, or scores, on a given hypothetical parameter. Thus, one may distinguish between the concept of "modes of measurement" and the hypothetical parameters upon which these measurement techniques are intended to produce scores. The lack of equivalence between particular measurement or scaling procedures and the hypothetical parameters upon which score values are desired may be illustrated by considering a situation where certain modes of measurement are inadequate for the task at hand. For instance, we have previously mentioned that the normative measurement orientation does not consider intra-individual variations in subdomains of normative traits. Similiarly, the pure ipsative approach does not attempt to work with normative variations between individuals. However, if one accepts the notion that a given set of performance scores may reflect both normative and ipsative sources of variation, then both of the above measurement orientations become inadequate as procedures for assigning appropriate normative and ipsative scale values to the data. Rather, some method which takes into account both sources of variation should be employed. Thus, the hypothetical parameters on which the data are to be given scale values are seen to be conceptually independent or superordinate to the particular manner in which the various marks on the parameters are developed. Hypothetical normative and ipsative behavioral parameters, then, represent two distinct and separate areas of concern for theories of behavior. It is not surprising that theories of psychology have progressed down both of these avenues of inquiry.

NORMATIVE AND IPSATIVE PSYCHOLOGICAL THEORIES

Modern psychology includes both normatively and ipsatively oriented theories. For instance, Krech's (1949) theory that differences in intelligence are due to different rates of cortical conductivity seems, essentially, a normative theory. The dimensions involved in Sheldon's (1944) body types also seem to be true normative parameters, as do the various traits of personality with which Allport (1937) has been concerned.

However, most theories concerning the mechanisms of psychological functioning assume that behavior is a function of relationships between psychological events occurring within a given individual. For instance, Helson's (1948) Adaptation Level Theory suggests that a perceptual response is determined by the relationship of the present stimulus to previously experienced stimuli. Similarly, Werner and Wapner's (1952) Sensory-Tonic Theory stresses the role of the relationship of sensory to tonic factors in determining perception. Thus, the above perceptual theories, many learning theories (Hull, 1943; Thorndike, 1932), and general personality theories (Freud, 1935; Lewin, 1935) all emphasize the importance of relationships between psychic events in determining behavior. These theories, then, are essentially ipsative in nature. Of course, many theories may include both normative and ipsative considerations. For instance, while Freud (1935) is primarily concerned with the relationships between psychic institutions, he also postulates normative, between individual differences in endowments of libidinal energy. Conversely, while Spearman (1927) is perhaps best known for the normative aspects of his theory, i.e., differences in intelligence are due to different amounts of mental energy, his theory also includes formulations about ipsative relationships between subfunctions of intelligence. These theories, and quite likely others, then, contain both normative and ipsative formulations.

It should also be noted that the normative-ipsative distinction is not represented in the language of individual differences. Individual differences are usually thought of only in a normative sense, i.e., in Anastasi and Foley (1949) differences between individuals are thought of as "a matter of degree. It is in this sense that individual differences are said to be quantitative rather than qualitative" (p. 59). However, this orientation tends to obscure the possibility that certain between-individual differences may be a function of varying ipsative, within-individual relationships, e.g., a psychoanalyst might argue that the differences between psychotics and normals are due to different intrapsychic relationships between ego and id. On the other hand, individual differences stemming from differing ipsative relationships can also be placed on normative scales, e.g., the above analyst might also say that one patient is more or less psychotic than another. Terminology is needed, then, to distinguish those individual differences which are due solely to variations on normative parameters from individual differences stemming from differing ipsative relationships. Hence, it is proposed that the purely normative individual differences be termed "first order" normative differences, while those individual differences stemming from ipsative, within-individual relationships might be termed "second order" normative differences.

It seems, then, that there are two legitimate, but independent realms of inquiry for a general psychology of individual differences. These are (*a*) the establishment of those normative parameters of differences within the species that are not merely phenotypical expressions of ipsative internal relationships, and (*b*) the establishment of ipsative parameters of internal organization. As already demonstrated empirically, the generalizations that hold at one level of analysis can differ from those that hold at the other level (Broverman, 1960, 1961; Mandler, 1959).

NORMATIVE AND IPSATIVE MEASUREMENT IN STATISTICS

Relatively little explicit attention is currently given to the distinction between normative and ipsative measurement.[2] Consequently, numerous inconsistencies in the applications of these orientations exist. For instance, most psychologists seem to prefer normative approaches to research problems even though many, if not most, of present day psychological theories lean heavily on ipsative conceptualizations. The development of statistical procedures for separating these different aspects of behavioral data has also been uneven. Some statistical techniques discriminate between normative and ipsative aspects of data while others do not. The following two hypothetical examples, involving typical research problems and statistical procedures, illustrate (*a*) an appropriate separation of normative from ipsative aspects of data, and (*b*) a confounding of these two types of information.

The first example involves assessing the significance of the difference between two means. An investigator wishes to determine whether motivating instructions improve performances on a particular task. Hence, the experimenter administers the task under a motivated and a nonmotivated condition to the same group of subjects. Assuming sequence effects are controlled, the experimenter might proceed to assess the significance of the difference between condition means via an analysis of variance technique. Since the experimental design employs repeated measurements, the total variance would first be partitioned into "between" and "within" individual sources of variation. The within individual variance would then be further divided into that associated with variation in conditions, and that which is not. After division by the appropriate degrees of freedom, the ratio of the two portions of within-individual variance would constitute an F test. If this test attains the required level of significance the experimenter is able to say that "The mean performance under the motivated condition is significantly different from the mean performance under the nonmotivated condition."

This statistical procedure clearly separates the normative from the ipsative aspects of the data. The between-individual variance reflects normative interindividual differences in task ability, independent of conditions, and is appropriately discarded since this variance is not the point of inquiry. The experimenter is not concerned with ability differences between subjects. He wishes only to know the effects of conditions on the subjects after such ability differences are controlled. The within-individual variance, which reflects ipsative or intraindividual variation, on the other hand, reveals that the experimental conditions had significant effects upon performance when these effects were measured against each subject's own norm. This statistical procedure, then, sensitively discriminates between normative and ipsative aspects of data.

---

[2] An outstanding exception is Cattell's (1944) paper concerning the use of normative and ipsative scores in psychology.

In contrast, correlational analyses typically give little attention to the normative-ipsative distinctions in behavior. Consequently, correlational procedures tend to confound interindividual with intraindividual variations in performance scores as illustrated in the following example:

The experimenter wishes to know the correlations between several tasks. Accordingly, he administers each task to each subject in his subject sample and correlates the various scores.

Table 1 presents hypothetical standardized scores for 12 subjects on four tasks. Table 2 shows that the Pearson product-moment intercorrelations of these tasks are positive, ranging from .69 to .93. This standard correlational procedure accepts these correlations as the relationships between test performances, which, indeed, they are. However, in the light of the previous discussion, we know recognize that these relationships between test performances are a function of two sources of variation: (*a*) normative differences between individuals in general ability levels, and (*b*) within-individual ipsative variations in abilities on different tasks. The correlation coefficients in Table 2, then, are confounded resultants of these two sources of variation. This is the same confounding that the experimenter took pains to avoid when assessing the difference between conditions means. The analogue of the mean difference technique in correlational procedures would be to eliminate normative differences in general ability level between individuals before examining the intraindividual variations that occur as a function of differences in task requirements. The more precise experimental question would then be "What are the relationships between abilities on different tasks when differences in general ability levels between subjects are held constant or statistically controlled?" As will be shown below, these relationships may be quite dissimilar to those obtained from the usual confounded method.

TABLE 1

HYPOTHETICAL PERFORMANCE DATA STANDARDIZED SCORES

Subject	Test 1	Test 2	Test 3	Test 4	Mean standard score
1	67	62	58	55	60.5
2	62	67	55	58	60.5
3	58	55	62	67	60.5
4	55	58	67	62	60.5
5	53	51	49	47	50.0
6	51	53	47	49	50.0
7	49	47	51	53	50.0
8	47	49	53	51	50.0
9	45	42	38	33	39.5
10	42	45	33	38	39.5
11	38	33	42	45	39.5
12	33	38	45	42	39.5

TABLE 2

INTERCORRELATION OF TESTS

Test	Test 2	Test 3	Test 4
Test 1	93	69	69
Test 2		69	69
Test 3			93
Test 4			

Note.—Decimals have been omitted.

Our problem is to separate the variance attributable to normative differences between subjects from that due to within-individual variations in ability to perform different tasks.

Normative variance between subjects occurs when some subjects tend to be generally above the group mean, while others tend to be below this reference norm. However, if the task per-

TABLE 3

HYPOTHETICAL PERFORMANCE DATA TRANSFORMED INTO IPSATIVE SCORES
$(X_i - \bar{X}_i)$

Subject	Test 1	Test 2	Test 3	Test 4
1	6.5	1.5	−2.5	−5.5
2	1.5	6.5	−5.5	−2.5
3	−2.5	−5.5	1.5	6.5
4	−5.5	−2.5	6.5	1.5
5	3.0	1.0	−1.0	−3.0
6	1.0	3.0	−3.0	−1.0
7	−1.0	−3.0	1.0	3.0
8	−3.0	−1.0	3.0	1.0
9	5.5	2.5	−1.5	−6.5
10	2.5	5.5	−6.5	−1.5
11	−1.5	−6.5	2.5	5.5
12	−6.5	−1.5	5.5	2.5

formances of each subject are expressed as ipsative scores, or deviations from each subject's mean level of performance, then all subjects will have the same status with respect to the new reference norms, i.e., the sum of the deviations above each ipsative norm will equal the sum of the deviations below for each subject. Hence, normative differences between subjects are, in effect, removed via ipsatization of performance scores. Correlation of the ipsative scores will then reveal the intrapsychic relationships between abilities on different tasks, free from the influence of differences in general ability between subjects.

The mean standard score of each subject is shown in Table 1. Table 3 shows the ipsative scores $(X_i - \bar{X}_i)$ of each subject on each task. Table 4 shows the intercorrelations of these ipsative scores. It will be noticed that some of the correlations are now negative, others positive. Thus, controlling normative interindividual differences in general ability levels considerably alters the correlational relationships between tasks, and, of course, any conclusions that might be drawn from these data. Instances where such controls have actually affected the results of studies have been reported by Mandler (1959) and Broverman (1960, 1961).

It seems clear then, that correlational analysis, a classical research tool in systematic psychology, has suffered from a lack of sensitivity to the normative-ipsative distinctions in behavior. Standard *R* factor analysis technique, which is an elaboration of the correlational method, suffers in the same manner, i.e., the factors extracted by these techniques are confounded expressions of normative and ipsative variation. The confounding in *R* factor analysis can be overcome in the same manner as illustrated above, i.e., by ipsatizing the various performance scores before starting correlational and factoring procedures. As in the above correlational example, factors derived from ipsative scores will frequently have large negative loadings or correlations with tests. A discussion of the numerical mechanics underlying this phenomenon may be found in Broverman (1961). That paper also pointed out that an ability factor defined by both positive and negative loadings can be viewed as a "choice point" in development since at some point in each individual's development a decision must be made as to whether he is to specialize in one or the other of the two classes of behaviors defining the factor. Other investigators (Cattell & Baggaley, 1956; Sweney, 1959) have interpreted this specialization or directional

TABLE 4

INTERCORRELATIONS OF IPSATIVE SCORES

Test	Test 2	Test 3	Test 4
Test 1	58	−79	−79
Test 2		−79	−79
Test 3			58

Note.—Decimals have been omitted.

aspect of ipsative scores as an index of motive strength.

RELATIONSHIP BETWEEN NORMATIVE PARAMETERS AND PARAMETERS DERIVED FROM DOUBLE-CENTERED IPSATIVE SCORES

The relationship between dimensions derived from ipsative scores distributed about a normative base-line, and the normative dimension itself is worthy of particular note. Such ipsative dimensions are internal subdivisions or differentiations of the larger normative dimension. Again, Spearman's (1927) theory of intelligence illustrates this point. Spearman argues that intelligence varies in units of mental energy. Normatively, then, the subjects differ from each other in terms of the amount of mental energy possessed by each. However, within intelligence, as mentioned earlier, Spearman posits subfunctions of eduction, reproduction, etc., the strength of which are also measured in terms of mental energy and which vary ipsatively within each individual. These subfunctions, then, are internal differentiations within the larger normative trait of intelligence.

This same part-whole relationship between ipsative and normative factors is obtained when the ipsative dimensions are empirically derived from double-centered ipsative scores via factor analysis. For instance, a subject's average rank performance for a battery of cognitive tests might be interpreted, normatively, as an indication of his general level of intelligence. On the other hand, the factors emerging from an analysis of the ipsative scores distributed about the average rank standings of subjects would represent internal differentiations of that normative trait.

This part-whole relationship of ipsative to normative factors may be contrasted with the relationships between factors produced by the usual normative *R* factor analysis procedures. Factors produced by this latter method are discrete, and juxtaposed externally to each other. Even though normative *R* factors may be correlated with each other, one normative factor is not thought of as a subdivision of another, as in the double-centered ipsative approach.

As pointed out by Stephenson (1953), factor analysis is typically used in an exploratory way to determine the various interdependencies between psychological functions or behaviors. When statistical procedures are used in this exploratory manner, one tends to allow the model embodied in the statistical procedures to guide one's conception of the organization of the behavior reflected in the data. Thus, the 30 years of research employing normative *R* factor analysis, with its model of externally juxtaposed factors, has left its mark on present conceptions of mental functioning, e.g., Burt's *The Factors of the Mind* (1946), Thurstone's *Primary Mental Abilities* (1938). While the part-whole model embodied in double-centered ipsative measurement is not new on the psychological scene—witness Spearman's (1927) theory of intelligence—a clearer recognition of its properties may exert a correspondingly greater influence upon our conceptions of the organization of behavior.

DETERMINATION OF APPROPRIATE NORMATIVE BASE-LINES

A major problem in double-centered ipsative measurement is the identification of appropriate normative baselines from which to develop ipsative scores. So far, we have assumed that a single normative base-line may be used for a given matrix of data. This assumption is implicit in the analysis of variance repeated measurements design. However, it is possible that more than one normative variable, hence, more than one base-line, is actually embedded in a given matrix of data. In this case, certain behaviors ought to be ipsatively measured against one base-line, others against another. Hence, a method for objectively establishing the presence of normative variables is needed. This problem may be resolved via the following considerations:

Any single performance score may be thought of as a resultant of normative and ipsative sources of variation. Consider, now, a situation where several tests all reflect a single normative trait, but some of these tests also reflect one ipsative dimension, other of these tests reflect a second ipsative dimension, still others reflect a third ipsative dimension, etc. Each test score of a given subject, then, consists of a constant representing his status on the normative dimension and a variable representing the standing of the task on a particular ipsative dimension. Since this normative constant varies between subjects, it tends to produce positive relationships between tests. The ipsative components of the various performance scores, on the other hand, will either correlate positively when they reflect the same ipsative dimension, or not correlate when they reflect different ipsative dimensions. Still other ipsative components may be inversely related and work towards negative relationships. Thus the normative components tending to produce positive relationships between tests will summate with those positive ipsative covariations, but will be weakened or canceled out by those independent and negatively varying ipsative components. The first order (simple structure) factors extracted from performance scores, then, will be those tests whose ipsative components have positive relationships augmented by their positively covarying normative components.

Since each of these factors have the same normative components, only the variation of the ipsative components can be acting to produce the different factors. Also, since each factor has the same normative component, these factors should correlate with each other. Accordingly, the factor intercorrelations ought to produce a second-order factor reflecting the presence of the normative trait. Of course, if the data actually contained more than one normative trait, then more than one second-order factor would be obtained. The number of second-order factors obtained from a factorial solution of performance data, then, indicates the number of normative dimensions present in the data. In turn, each normative dimension calls for its own base-line from which to assess ipsative variations of the tests included in the second-order factor. Thus, the mean group standing of all tests included in each second-order factor would be obtained for each subject. Deviations of individual tests from these means would then reflect ipsative variations within a given normative dimension. Other methods of obtaining appropriate base-lines for ipsative oriented research problems may be found in Broverman (1960).

CONCLUSION

It may be well to conclude by attempting to see the issues discussed in this paper in a broader perspective. It seems that the measurement of behavior is more complicated than previously suspected. One wonders why this should be so; why one can not proceed to a more straightforward measurement of behavior. The reasons seem at least three-fold:

First, human behavior is so bewilderingly varied and multifaceted that few psychologists are willing to stay at the level of raw, uncategorized, observations of behavior. Such an enterprise would produce an infinity of parameters, variables, and scales. Rather, one typically seeks to reduce a large number of different behaviors to a smaller number of common underlying parameters. The pertinent variables of behavior, then, are not accessible for direct measurement. One must fractionate, peel off, and discard the surface aspects of different behaviors in order to uncover and obtain an estimate of the underlying variables. Thus, in measuring behavior, phenotypical surface appearances are discarded in favor of inferred genotypical variables.

A second aspect of human behavior which contributes to the complexity of its measurement is that any given behavior can simultaneously represent more than one underlying variable. A given behavioral act may be a composite expression of many different variables. Thus, one must separate not only phenotype from genotype, but also genotype from genotype.

Finally, we must recognize that all human behavior potentially participates in two different realms of human variation, i.e., species and individual. Each behavioral act is, potentially at least, a simultaneous expression of the human species' as well as of the individual's repertoire of variation. It is this dual aspect of behavior which produces the particular normative-ipsative model elaborated in this paper. Failure to represent this fact in measurement and statistical operations must lead to a confounding of these two different aspects of human existence and variation. If one accepts the proposition that the laws or regularities which characterize variations in behavior between individuals are not the same as those which characterize variations in behavior within individuals, then such confounding is to be studiously avoided. Hence, an orderly inquiry into the nature of human behavior requires a stringent separation of the between-individual species parameters from the ipsative intraindividual organizational aspects of behavior in order that the separate relational networks of each realm may be established. While this may involve considerable more work than previous methods, the results may be proportionately more rewarding.

REFERENCES

ANASTASI, ANNE, & FOLEY, J. P. *Differential psychology.* New York: Macmillan, 1949.

ALLPORT, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

BROVERMAN, D. M. Cognitive style and intra-individual variations in abilities. *J. Pers.*, 1960, **28**, 240–256.

BROVERMAN, D. M. Effects of score transformations in *Q* and *R* factor analysis techniques. *Psychol. Rev.*, 1961, **68**, 68–80.

BURT, C. L. *The factors of the mind.* London: London Univer. Press, 1946.

CATTELL, R. B. Psychological measurement: Ipsative, normative and interactive. *Psychol. Rev.*, 1944, **51**, 292–303.

CATTELL, R. B., & BAGGALEY, A. R. The objective measurement of attitude motivation: Development and evaluation of principles and devices. *J. Pers.*, 1956, **24**, 401–423.

FREUD, S. *A general introduction to psychoanalysis.* New York: Liveright, 1935.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HULL, C. L. *Principles of behavior: An introduction to behavior theory.* New York: Appleton-Century, 1943.

KRECH, D. Notes toward a psychological theory. *J. Pers.*, 1949, **18**, 66–87.

LEWIN, K. *A dynamic theory of personality: Selected papers.* New York: McGraw-Hill, 1935.

MANDLER, G. Stimulus variables and subject variables: A caution. *Psychol. Rev.*, 1959, **66**, 145–149.

SHELDON, W. H. Constitutional factors in personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944.

SPEARMAN, C. *The nature of intelligence and the principles of cognition.* London: Macmillan, 1927.

SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In, *Kentucky Symposium: Learning theory, personality theory, and clinical research.* New York: Wiley, 1954. Ch. 1.

STEPHENSON, W. *The study of behavior.* Chicago: Chicago Univer. Press, 1953.

SWENEY, A. B. The objective measurement of motivation in children. *Amer. Psychologist,* 1959, **14**, 355. (Abstract)

THORNDIKE, E. L. *The fundamentals of learning.* New York: Teachers College, Columbia University, 1932.

THURSTONE, L. L. *Primary mental abilities.* Chicago: Chicago Univer. Press, 1938.

WERNER, H., & WAPNER, S. Toward a general theory of perception. *Psychol. Rev.,* 1952, **59**, 324–338.

(Received October 24, 1960)

Psychological Review
1962, Vol. 69, No. 4, 306–328

FRUSTRATIVE NONREWARD IN PARTIAL REINFORCEMENT AND DISCRIMINATION LEARNING:

SOME RECENT HISTORY AND A THEORETICAL EXTENSION[1]

ABRAM AMSEL

University of Toronto

The overall objective of this paper is to examine the current status of a concept of frustrative nonreward in behavior theory. The examination takes two forms: a general discussion and survey of recent work on motivating and inhibiting properties of nonreward, followed by a theoretical extension of frustrative-nonreward theory to certain prediscrimination phenomena.

In the first part of the paper a brief historical account will be presented of recent changes toward an "active" conceptualization of nonreward, particularly in neo-Hullian theory, the current status of such an active nonreward concept in explanations of partial reinforcement effects will be set in a broader framework of partial reinforcement theories, and some recent data will be presented in support of the idea that discrimination learning involves frustrative-nonreward effects.

The second more theoretical portion of this paper stems from an earlier frustrative conceptualization of nonreward (Amsel, 1958). It represents an attempt to extend frustrative-nonreward theory to situations in which discrimination learning is preceded by various amounts and kinds of "prior experience." These prior experiences will be termed "prediscrimination treatments" and will involve training to approach in the presence of one or both (separately) of the eventual discriminanda, under conditions of partial or of continuous reinforcement.

HISTORY

Hull's adoption of what he termed the Mowrer-Miller hypothesis as a basis for a two-factor theory of inhibition (Hull, 1943) was, at the same time, the acceptance of a "passive" conception of nonreward and rejection of a more active conception which he and others had held. Certainly, Pavlov's nonreinforcement, producing internal inhibition, was an active conception. Spence's (1936, 1937) theory of discrimination learning, clearly in the Pavlovian tradition, also assigned special inhibitory properties to nonreinforcement. In some of his more recent systematic writing, Spence (1956) points out that his position on the Mowrer-Miller hypothesis was never in agreement with Hull's.

Emotional-motivational effects of nonreward, as they relate to learning, were described in an article by Miller and Stevenson (1956), which attributed the agitated behavior of rats during extinction to effects of nonreinforcement carried over from one trial to the next. This is an example of an active motivational conception of nonreward. A similar emotional-motivational interpretation of nonreward can be found in Skinner (1938).

In Hull's 1943 theory, which dominated the psychology of learning in the forties, nonreinforcement was an event without direct inhibitory or motivational effect as it was also in the early mathematical-model treatments of learning (Bush & Mosteller, 1951; Estes, 1950). By about 1950, however, expressions of dissatisfaction with the passive theoretical status of nonreward were emerging. These expressions were in theoretically oriented experiments and in theoretical papers. Most of the early defections from 1943 Hull, in this connection, were by persons close to the Hullian point of view. There are suggestions that Hull, himself, was getting ready to change his position on this shortly before his death (Hull, 1952, Ch. 5).

---

[1] A shorter version of this paper was presented to the Psychology Colloquium of the University of Illinois in May 1961. The preparation of this paper and much of the experimental research to which it refers were supported by grants (G-5527 and G-13895) from the National Science Foundation.

Rohrer's (1949) treatment of extinction in terms of a "frustration drive" was one of the first expressions of a desire to give nonreinforcement a more active role in inhibition.[2] However, while Rohrer's reference to frustration was specifically related to the operation of nonreinforcement, it provided no particular mechanism for frustrative inhibition. At about the same time, the well-known experiments of Virginia Sheffield (1949, 1950) appeared, relating partial reinforcement acquisition and extinction effects to the spacing of trials and containing suggestions of both frustration drive and frustrative inhibition. These articles and another by Stanley (1952), along with a late experimental paper by Hull and others (Hull, Livingston, Rouse, & Barker, 1951) and Hull's (1952) last book, all contained at least the germ of a change toward a more active conception of nonreinforcement. An experimental paper by Denny and Dunham (1951) offered inhibition based on nonreinforcement (frustration) as an alternative to reactive inhibition to account for differential nonreinforcement effects in a T maze.

Brown and Farber (1951) outlined a theory in which nonreward was one of several antecedents to frustration. In this paper, the major emphasis was on the definition of frustration in terms of competing response tendencies and the authors were concerned only with the motivational (drive) properties of frustration, and not with inhibitory effects. In the same year, a paper by Amsel (1951) made the point that fractional anticipatory frustration (r_F), which is the classically conditioned form of the frustrative reaction to nonreward, provides a mechanism for conceptualizing the active properties of nonreinforcement and should be regarded as a determiner of inhibitory effects. And, at about the same time, Seward (1951) published a theoretical paper which mentioned, without elaboration, the possible functional properties of an anticipatory frustration factor.

The remainder of the decade, 1950 to 1960, has seen a number of additional experimental and theoretical attacks on the problem of nonreward. On the one hand there are experiments and theoretical treatments by investigators in the Hullian tradition of behavior theory which have sought to extend and revise Hull's theory in regard to the status of nonreward. The other hand holds a number of published studies representing a variety of theoretical and nontheoretical interests which have contributed information about frustrative nonreward.

---

[2] It is questionable that "inhibition" is the appropriate and correct term to use in regard to response decrement related to frustrative nonreward. Better terms would be "interference" or "competition," since frustrative inhibition cannot be regarded as directly affecting the excitatory tendency out of which it grows. It is rather a new excitatory tendency which competes with the older one.

Much of this work has been reviewed earlier (Amsel, 1958; Lawson & Marx, 1958); but some of it is very recent. The experiments fall into three categories: (*a*) demonstrations of the motivating effects of frustrative nonreward, i.e., the carried-over, enhancing effect of nonreward on the vigor of immediately following behavior. This has been termed the Frustration Effect (FE). Several studies, by far the greatest number, have been directed not only at demonstrating the effect but also at examining the variables determining its occurrence and strength (Amsel, Ernhart, & Galbrecht, 1961; Amsel & Hancock, 1957; Amsel & Penick, 1962; Amsel & Roussel, 1952; Bower & Stocks, 1960; Longstreth, 1960; Marzocco, 1951; Penney, 1960; Roussel, 1952; Seward, Pereboom, Butler, & Jones, 1957; Wagner, 1959). (*b*) The second category consists of experiments which are mainly concerned with the frustration drive stimulus, but which also touch on frustration reduction as reinforcement (Amsel & Prouty, 1959; Amsel & Ward, 1954; Tyler, Marx, & Collier, 1959). (*c*) And finally, there are a few studies which deal with the inhibitory properties of frustrative nonreward. The inhibitory mechanism is fractional anticipatory frustration (r_F), a classically conditioned form of the goal response to frustrative nonreward. The mechanism and its theoretical properties have been identified in detail (Amsel, 1958; Spence, 1960), and experiments testing implications of such a mechanism are available (Bower, 1961; Goodrich, 1959; Haggard, 1959; Wagner, 1961a, 1961b). Of particular importance in this connection are a series of experiments by Wagner (1961b) which go far toward establishing the proposition that frustrative nonreward has many of the properties of punishment, and that r_F operates in many respects like fear. These experiments would support such statements as (*a*) the greater the strength of frustration the faster the extinction (avoidance) of a nonrewarded response; (*b*) cues paired with frustrative nonreward acquire motivational properties; and (*c*) the cessation of cues previously paired with frustrative nonreward serves to reinforce a new response.

There are several other experiments on nonreward (or reward prevention) as a variable in learning which provide support for nonreward as a factor in inhibition. For example, studies by Lambert and Solomon (1952) and by Adelman and Rosenbaum (1954) related extinction to the blocking of instrumental behavior at various distances from the goal. An experiment by Adelman and Maatsch (1955) found that resistance to extinction depends to some extent on what the subject is allowed to do when it finds no reward in the goal box, i.e., is frustrated. The suggestion is that resistance to extinction is high when the subject makes a response to frustration which removes it from the frustrating situation rather than one which keeps it in the apparatus. Ferster (1957, 1958) found that stimuli signaling "time out" from positive reinforcement acquire aversive properties. He also showed that the withdrawal of a positive conditioned reinforcer had the functional properties of punishment, suppressing the rate of responding much like the presence of a negative reinforcer. These experiments point clearly to the inhibitory (aversive) properties of nonreward.

Some recent experiments, which show decreasing resistance to extinction with increasing numbers and magnitudes of reinforcement, must be regarded as supporting a conception of nonreward as an active factor. For example, Hulse (1958) and Armus (1959) found that larger magnitudes of reinforcement in acquisition are followed by faster extinction. Very recent reports by Reynolds and Siegel (1961), by Pavlik (1961), and by Wagner (1961a) were to the same effect. North and Stimmel (1960) have shown that rats reinforced 90 or 135 times extinguish faster than those reinforced 45 times. Similar results have been reported by Reynolds, Richter, and Carlock (1960). They found faster extinction following .5- than following .1-gram reward, and faster extinction after 66 acquisition trials than after 6, 12, or 30 trials. This kind of result is quite in accord with results from studies of overlearning and discrimination reversal, which we will look at later. Together, they are as incompatible with early Hullian theory, or any other theory requiring monotonicity and positive relationships in these variables (e.g., the various mathematical approaches), as were the early partial reinforcement findings. We will come back to some of these experiments later, and my discussion of partial reinforcement effects will show why a theory which includes a frustrative nonreward hypothesis must predict such results.

There are by now several experiments which can be taken to indicate that nonreinforcement is more important than reinforcement in discrimination. Chronologically, these are studies by Fitzwater (1952), Grove and Eninger (1952), Eninger (1953), Shoemaker (1953), Cantor and Spiker (1954), Grice and Goldman (1955), Birch (1955), and Solomon (1956), to mention some of the earlier ones. A very recent study which makes this same point has been reported by Lachman (1961). This is not the place for a detailed description of these studies; however, the burden of evidence is that avoidance of nonreward is a more powerful factor in discrimination than approach to reward; and this is, of course, very relevant to a point of view which makes nonreward a determiner of inhibition. Perhaps Harlow and Hicks (1957) were reacting not only to their own experimental findings but also to the weight of this kind of evidence when they argued for a "uniprocess" rather than a "duoprocess" theory of discrimination learning, the single process being inhibition. The weakness of their argument is that the negative process (r_F, in my terms) may well depend on the prior existence of a positive process (see Amsel, 1958; Spence, 1936, 1937), and one is back to two processes again.

Other examples of a return to an active conception of nonreward outside of the Hull-Spence camp can be found in the recent work of Mowrer and Estes. Mowrer, in his *Learning Theory and Behavior,* makes frequent use of the notion of nongoal events arousing "frustration" or "anger," in addition to "disappointment" and expresses agreement with the position I favor in regard to frustrative goal events (Mowrer, 1960, p. 409). Estes (1959) has recently acknowledged that statistical learning theory must cope with the problem of an active negative process based on nonreward in these words:

> . . . I have found that in the naive animal an unreinforced trial produces no apparent change in response probabilities, but that after a series of reinforced runs to a given side of a T maze, the introduction of nonreinforcement yields a decrement in probability of response to that side. . . . Apparently under partial reinforcement schedules, the effect of nonreinforcement varies between these extremes . . ." (p. 417).

From this brief sketch of the recent history of the concept of nonreward, we move to a description of the role of active conceptions of nonreward in interpretations of noncontinuous (partial) reinforcement effects.

NONREWARD IN INTERPRETATIONS OF PARTIAL REINFORCEMENT

In his 1958 chapter on learning, Lawrence differentiates between "intertrial" and "intratrial" interpretations of the Partial Reinforcement Effect (PRE)—the finding that partial or intermittent reinforcement in learning leads to greater resistance to extinction than does continuous reinforcement. Intertrial explanations attribute the PRE in one way or another to immediately carried-over traces of stimulation from one trial to the next in acquisition. According to such interpretations, it is on the basis of such traces that acquisition is more clearly discriminated from extinction by continuously than by partially reinforced subjects. Intratrial explanations ascribe the PRE to some kind of learned mechanism developing on each trial in acquisition and having its effect, associatively, later in extinction. There have been two classes of intertrial explanations. In both cases the classes can be termed Cognitive-Expectancy (C-E) interpretations, and Stimulus-Response (S-R) interpretations.

Generally speaking, the intertrial interpretations were an earlier development than the intratrial interpretations. Early C-E interpretations were the "common sense" expectancy hypothesis which emerged out of the early important experiments of Humphreys (e.g., 1939, 1940) and the discrimination hypothesis developed and tested by Bitterman and his associates (e.g., Bitterman, Fedderson, & Tyler, 1953). On the S-R side, the major intertrial explanation (apart from the "response unit" interpretation which applies to chained or highly massed responses but not to discrete trial learning) was the Hull-Sheffield hypothesis tested by Sheffield (1949). Weinstock's (1954) demonstration that partial reinforcement effects were not eliminated even when trials were separated by 24 hours, and an experiment by Wilson, Weiss, and Amsel (1955) disconfirming the earlier results of Sheffield helped to set the stage for, and indeed to provide, intratrial explanations of, the PRE. Much later, in fact very recently, there has been strong contraindication for a discrimination hypothesis in the work of Jenkins (1961b) and of Theios 1962). Jenkins, using pigeons in a Skinner-type situation, and Theios, using rats in a runway, have found that the PRE is sustained when a large number of continuous reinforcements are interpolated between discrete trial partial reinforcement and extinction. This means that a large number of continuous reinforcements immediately before extinction and following partial reinforcement does not neutralize the partial reinforcement effect. Since the basis for discriminating acquisition from extinction is now the same for partially as for continuously reinforced subjects, and the PRE persists, an interpretation is required, as Theios indicates, in terms of "relatively permanent effects of nonreinforcement which can be sustained through blocks of continuous reinforcement."

Intratrial explanations of the PRE concern themselves in one way or another with processes following nonreward. Until recently, the intratrial explanations were the exclusive property of S-R psychologists. Weinstock's (1954) contiguity-interference hypothesis was perhaps the earliest of these, but was followed in short order by a variety of neo-Hullian hypotheses. All of these are what I would term conditioning-expectancy hypotheses as opposed to cognitive-expectancy hypotheses of the PRE. All depend upon constructs of the anticipatory goal response type, representing responses which develop in strength within trials. There are three neo-Hullian interpretations of the PRE, all of which were first proposed to account for particular experimental findings. The first emerged out of a study by Wilson, Weiss, and Amsel (1955) and was subsequently elaborated and tested in some detail (Amsel, 1958). It attributes the PRE to the evocation of frustration by nonreward in partial reinforcement, to the conditioning of anticipatory frustration ($r_F - s_F$) to cues in the instrumental sequence, and to the association of s_F with approach responses in acquisition. Logan's interpretation of the PRE (Logan, Beier, & Kincaid, 1956) is that it depends on the extinction of r_G to "postreinforcement-time cues." Kendler and his associates (Kendler, Pliskoff, D'Amato, & Katz, 1957) have offered an interpretation similar to the one I have favored, except they choose a neutral designation of nonreward effects in acquisition and refer to anticipatory nonreward and not to anticipatory frustration. Spence (1960) has recently come out in favor of the same kind of interpretation and has employed the r_F construct to account for acquisition phenomena in partial reinforcement. Consequently, it seems safe to say that a prominent interpretation of the PRE in neo-Hullian writings is a conditioning-expectancy position with identifies fractional anticipatory frustration as well as fractional anticipatory reward.

Recently, Festinger (1961) has called attention to the importance of "insufficient rewards" in a cognitive-expectancy (cognitive dissonance) interpretation of the PRE. Cognitive dissonance (Festinger, 1957) is a hypothesis which recognizes the active properties of nonreward. In its use of delay of reward procedures to enhance attractiveness of a goal in a delay of reward situation, Festinger's treatment of insufficient rewards is very similar in conceptualization to some earlier work of Olds (1953, 1956). On the basis of experiments with children, Olds concluded that the reward value of a secondary rewarding stimulus (S^r) could be increased—and hence the strength of approach responses leading to S^r could be increased—by delaying the presentation of S^r *after* overlearning. He showed that tokens which could ultimately be exchanged for more primary reward increased in reward value when their presentation was delayed, if these delay trials followed a number of trials in which their presentation was immediate. Olds suggested that such a procedure constitutes "practice at wanting" and that such practice at wanting increases the value of that which is wanted (S^r). Festinger would seem to be taking the same position in his animal research. In one of several experiments dealing with dissonance in rats, he employs a double runway in which a start box is separated from a mid box by an alley, and the mid box is separated from an end box by another alley. Subjects are delayed in the mid box in the manner of Holder, Marx, Holder, and Collier (1957) and Wagner (1959), and the finding is that the subjects delayed in the mid box, before being allowed to run to the end box and find food, will continue to run to the mid box longer on test trials than the subjects always fed in the mid box as well as in the end box. Festinger's (1961) interpretation is that the delayed subjects develop dissonance as a result of the delay, and that the "dissonance tends to be reduced by developing some extra preference about something in the situation. The existence of this extra preference leads to the stronger inclination to continue running during extinction trials" (p. 9).

The position of Festinger, like the earlier one of Olds, is clearly that nonreward (or delay of reward) is an active factor determining an increased want or attraction. However, there seems to be an important difference in what increases in attractiveness between Festinger's earlier treatments of dissonance in humans (Festinger, 1957) and his application of the concept to the explanation of the PRE in rats (Festinger, 1961). While specificity is not always a characteristic of cognitive theory, it does seem as though the 1957 treatment (and even parts of the 1961 paper) did specify that the violation of a cognitive expectation produces dissonance which is resolved by upgrading the value of the previous goal object. However, in the treatment of "insufficient rewards" in terms of the prototype experiment with start box, mid box, and end box, dissonance is said to develop in the mid box, where the subject never has found food; and dissonance is reduced "by developing a preference for something about the *place* where the dissonance was introduced . . . by developing some liking for the place where they were not rewarded . . ." (p. 10). This latter position is used to explain, in terms of dissonance, why the subjects never fed in the mid box run longer to the mid box on nonreward test trials than the subjects always fed there before. The same thinking is then carried over to explanation of the PRE. In terms of the S-R analysis, the subjects have been trained to approach the mid box of Festinger's situation despite anticipatory frustration (s_F) cues because, in the period prior to testing, they were always finally fed in the end box. In C-E terms the subjects developed an extra preference for the nonreward situation—the mid box in which they had never been rewarded. This would leave to be answered such questions as: Why does successive nonreward result in extinction? And, how do the subjects learn to discriminate cues leading to reward from cues leading to nonreward?

The S-R treatment of nonreward effects in partial reinforcement (Amsel, 1958; Kendler, Pliskoff, D'Amato, & Katz, 1957; Spence, 1960; Wilson, Weiss, & Amsel, 1955) does not suggest that nonreward (or delayed reward) increases "wanting" in Olds' terms, or "preference about something in the situation" in Festinger's. This type of neo-Hullian approach is a description of a mechanism operating through anticipatory nonreward which can lead to either rapid extinction of behavior or to increased resistance to extinction, and which can either facilitate discrimination learning or retard it. I will claim that neither increased "wanting" nor "preference"—in fact, no "common sense" cognitive interpretation—can account for all of the effects which nonreward seems to have.

The difference between the two positions, in the cognitive language of Olds and Festinger, is as follows: While the C-E position has nonreward *increasing* attractiveness of the nonreward situation, the S-R position implies a *decrease* in attractiveness. The partial reinforcement effect is, according to Festinger, due to increased attractiveness. According to my position it is the result of training to persist in behavior despite decreased attractiveness. In some other respects the positions are similar: both make nonreward the important determinant of the effect, and both stress the importance of developing (intratrial) associative tendencies during acquisition as determinants of ultimate resistance to extinction. Neither necessitates an assumption that the subject discriminates acquisition from extinction (Bitterman, Fedderson, & Tyler, 1953).

The details of the S-R position which I favor have been presented elsewhere (Amsel, 1958) in the form of a sequence of hypotheses identified with four stages of practice in discrete trial partial reinforcement or in discrimination learning (the first three stages are essentially the same for both). A fifth hypothesis which applies to discrimination learning has been added, and this will be important for some data which follow. The essence of the position is outlined below in relation to an apparatus we have been using. It was designed to differentiate the immediate motivating (energizing) effects of frustration exemplified by the so-called Frustration Effect (FE) from fractional anticipatory frustration (r_F), which represents the inhibiting feature of frustration. The apparatus is essentially, two runways in series, either in a straight line (e.g., Amsel & Roussel, 1952) or in the form of an L (e.g., Amsel & Hancock, 1957). The apparatus has a start box (SB), a first runway (Runway 1), a first goal box (G1), a second runway (Runway 2), and a second goal box (G2) in successive arrangement. Reward and nonreward are manipulated in G1, and the apparatus permits a separation of the frustrated response and changes in running speed indicating r_F (measured in Runway 1), from the frustration motivated response and changes indicating the FE (measured in Runway 2).

The sequence of hypotheses in terms of five stages of practice is as follows:

1. In Stage 1, $r_R - s_R$[3] is developing in Runway 1 with early rewards, and nonreward (in G1) has no particular effect, as measured by the FE in Runway 2.

2. After the development of $r_R - s_R$ for the Runway 1 response, nonrewards elicit frustration, measured by the FE in Runway 2. This is Stage 2.

3. When nonrewards elicit frustration, the cues in Runway 1 previously evoking r_R now also begin to evoke r_F. In Stage 3 these antedating goal response tendencies are temporarily in competition.

4. Since r_R and r_F cannot be elicited separately by differential cues in partial reinforcement, as they can in the latter stages of discrimination learning, and since the temporary conflict in partial reward training is resolved in favor of running to the intermittently rewarding goal box, s_F becomes associated with the instrumental approach response in Stage 4 of partial reinforcement training, providing the mechanism for the partial reinforcement effect. When extinction is carried out, partially reinforced subjects have been trained to respond (approach in the presence of antedating,

---

[3] I employ the term, $r_R - s_R$, in the same manner as it is used in an earlier development, and for the same reason: ". . . $r_G - s_G$ is a general term covering all types of antedating conditioned response. Separately, these might be designated $r_R - s_R$, fractional anticipatory reward; $r_P - s_P$, fractional anticipatory punishment, usually termed fear or anxiety; and $r_F - s_F$, fractional anticipatory frustration" (Amsel, 1958, p. 102). The meaning of $r_R - s_R$ is, however, not different from $r_G - s_G$ as Spence (1960) uses it in a similar connection. All that is intended here is a clarification of terminology in light of the fact that goal events may be positive (such as to elicit approach) or negative (such as to elicit avoidance).

frustration produced stimuli), whereas consistently reinforced subjects have not. In discrimination situations there is a basis for the differential elicitation of approach and avoidance by s_R and s_F respectively in Stage 4.

5. Following discrimination, when r_F consistently precedes nonreward, the FE following nonreward should grow weaker, since nonreward is defined as frustrating only in the presence of r_R.

I have been suggesting that in a very important area of the analysis of behavior, the explanation of persistence of behavior in the face of nonrewards attributable to partial reinforcement, the role of nonreward as an active process has emerged in both cognitive-expectancy and in S-R conditioning interpretations. It would certainly seem to me that this is an area where the clearer, better worked out S-R language and conceptualization might provide a handle for a more workable expectancy theory.

There is, in addition, the matter of a basic difference between the cognitive and S-R interpretations of the PRE: the former says nonreward increases attractiveness; the latter says nonreward decreases attractiveness. Both positions can account for the PRE; both can account for faster extinction following large rewards than following smaller rewards; both can account for faster extinction following many rewards than following fewer rewards. However, a cognitive theory such as Festinger proposes would have difficulty with a finding of Hulse (1958) which has recently been confirmed by Wagner (1961a). In Hulse's words: ". . . large as compared with small rewards produced greater resistance to extinction if partial reinforcement were used, but *less* resistance to extinction if continuous reinforcement were used . . ." (p. 56). In the cognitive language of Festinger this would have to mean that the subjects develop "some extra preference about something in the situation" (greater resistance to extinction) when rewards are *small and continuous;* and that they do not develop the extra preference, in fact quite the opposite, when rewards are *small and partial.* The result of Hulse and of Wagner is clearly in line with the S-R interpretation in terms of anticipatory responses: Extinction is *faster* following continuous rewards of large than of small magnitudes because of the greater FE in extinction and subsequent faster development of r_F. However, in partial reinforcement there is opportunity to elicit FE and to build r_F during acquisition. Under these circumstances, the larger magnitude of reward permits the mechanism for the PRE—the connection of s_F to continued approach—to develop more strongly. Consequently extinction should be *slower* following partial rewards of large magnitude than following partial rewards of small magnitude.

I am claiming that the S-R conditioning analysis of nonreward effects has much to recommend it. The question is, How much? To break the question down: Does inclusion of the concepts r_R and r_F in the analysis of partial reinforcement and discrimination learning provide us with explanatory power which a more parsimonious approach-avoidance analysis (e.g., D'Amato & Jagoda, 1960; Nissen, 1950) does not? Does this type of analysis of nonreward effects provide predictive power which a cognitive treatment of nonreward (e.g. Festinger, 1961) does not?

My attempts to convince the reader that the answer to these last two questions is "yes" will take two forms: (*a*) Some data will be presented which bear on predictions already made on the basis of this type of analysis. I will claim that data such as these are not readily deducible from either a less specific cognitive language or from a more parsimonious approach-avoidance language. (*b*) I will offer an extension of frustrative nonreward theory designed to generate predictions about the course of discrimination learning following various prediscrimination experiences with partial and continuous reward.

Before proceeding to this extension of the theory of frustrative nonreward, a brief comment on the relation of the present position on nonreward to other theories of frustration. As I have indicated before (Amsel, 1958), ". . . we have been interested in certain active properties of nonreward following reward and no more than this is meant by the term frustration as we use it" (p. 103). Nevertheless, the mechanism which has been outlined to explain the PRE, particularly the conditioning of r_F and the connection of s_F to continued approach as a result of intermittent reward, is certainly a candidate for the learning theory counterpart of "frustration tolerance" as employed in social psychology and psychopathology. The treatment of frustration from the viewpoint of theories of personality is beyond the scope of this paper; and there is no intent here to grapple with the multiple meanings the concept has in those contexts. However, it would seem possible to recast portions of the frustration theory of Rosenzweig (e.g. 1944) which deal with nonreward (failure or delay of "gratification") into the present terminology; and the treatment of frustration by Rotter (1954), as a discrepancy between reward expectancy and its occurrence, seems compatible with the mechanisms hypothesized here and with the theoretical extension I will propose.

NONREWARD EFFECTS IN DISCRIMINATION LEARNING: SOME DATA

According to the sequence of five hypotheses on the role of nonreward and partial reinforcement in discrimination, it should be possible to show that the FE, measured in Runway 2 of a double runway, bears certain relationships to various stages of discrimination learning. More specifically, (*a*) evidence of nonreward related frustration, measured in Runway 2, should precede any evidence that the subjects are running faster to S+ and slower to S− in Runway 1, and (*b*) at some point following the appearance of discriminative behavior in Runway 1, a reduction in the magnitude of the FE measured in Runway 2 should become apparent.

In an unpublished doctoral dissertation, Ward (1961) employed the double runway to record discrimination to positive (e.g., black) and negative (e.g., white) cues presented singly in Runway 1, and the concurrent development of frustrative effects in Runway 2. A single runway condition (Runway 1 alone) was used to study the same discrimination uncontaminated by Runway 2 performance. There were 20 subjects in the double runway group and 10 subjects in the single runway comparison group. They were run six trials a day for 30 days, and the intertrial interval was about 10 minutes. Figure 1 presents data for the 20 subjects in the double runway. The top panel shows the developing FE in Runway 2. The bottom panel shows discrimination starting at about Day 7 (42 trials), and at a point well after the beginning of FE development on Day 3 (18 trials). There is an indication that the FE almost disappears (perhaps only temporarily) at a point where separa-

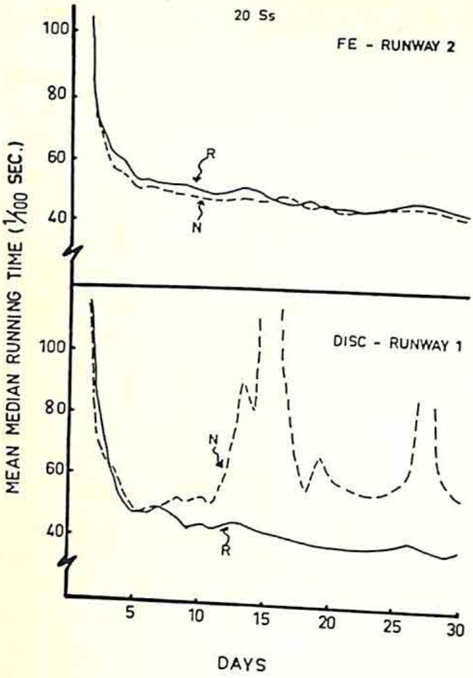


FIG. 1. The appearance and diminution of the frustration effect in Runway 2 (top panel) in sequential relationship to the development of the successive discrimination in Runway 1 (bottom panel). (From Ward, 1961)

tion of the discrimination curves is greatest. This occurs at about Day 15 (90 trials).

It could be argued the Ward result is an artifact: that different subjects reach criterion at different times. Anticipating this, I selected the data of seven of Ward's subjects which reached the discrimination criterion on the same block of trials. Figure 2 presents these data in 2-day blocks and shows essentially the same relationships as in Figure 1.

Figure 3 presents data from the single runway comparison group and shows the Runway 1 discrimination unaffected by any subsequent Runway 2 performance. Even in the absence of the second runway there is a tendency for the N and R curves of the discrimination to converge on later trials (see Berry, Verplanck, & Graham, 1943). These data of Ward, which have been duplicated in our laboratory as by-products of other studies, are taken as evidence for the involvement of frustrative effects in discrimination in a manner such as our analysis suggests.

In another experiment using the same double runway apparatus, Ward (1961) studied rate of discrimination learning in Runway 1 as related to number of prediscrimination rewards. In this experiment, which was similar to an earlier one of Shoemaker (1953) performed in a single runway, the response of running in Runway 1 and entering G_1 was rewarded 0, 12, or 48 times prior to learning a black-white discrimination in Runway 1. An implication of our position is that, with initial positive tendencies to the two discriminanda equal, rate of dis-

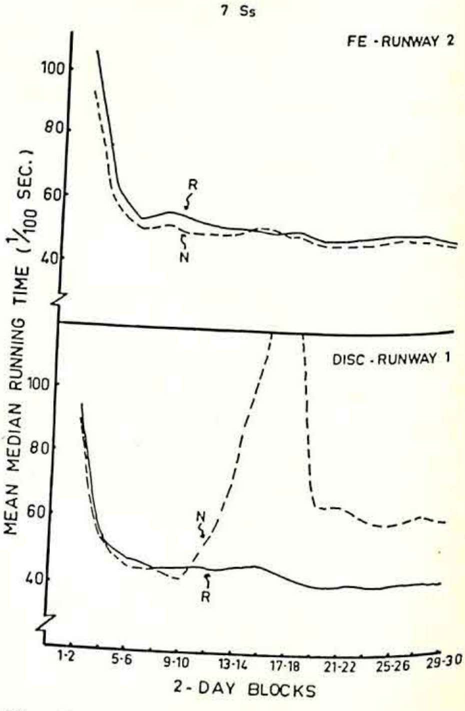


FIG. 2. These are the relationships shown in Figure 1 for seven subjects selected because they reached the criterion of discrimination on the same block of trials. (Adapted from Ward, 1961)

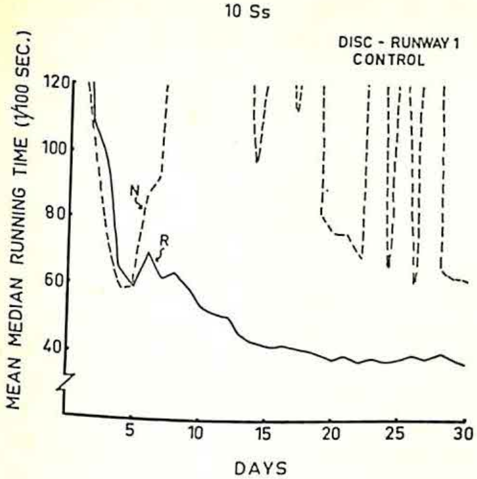


FIG. 3. The successive discrimination in a group which ran only in Runway 1. (Extreme N times going up to 3 or 4 seconds are not shown. Note the decline of the nonreward curve in the single runway group as in the double runway groups.)

crimination learning should be a positive function of the strength of r_R to the negative stimulus. This meant that prediscrimination trials were all rewarded in G_1 and that Runway 1 was black on half of these trials and white on the other half. On discrimination trials, Runway 1 was black when G_1 contained reward and white when G_1 contained no reward (or vice versa). Second runway performance was measured on every discrimination trial following reward or nonreward in G_1 to get a measure of the FE, whose strength at different stages could then be related to the developing discriminative performance. Table 1 shows the results of this experiment for both running time and goal entry measures for two criteria of learning. A more detailed analysis of the data in Ward's dissertation also indicates the appearance of the FE before discrimination and its diminution or disappearance following discrimination. These data are in accord with, and go a little beyond, previously cited studies which stress the importance of nonreward in discrimination.

A recent experiment by Jenkins (1961a) demonstrates the importance of nonreward effects in discrimination learning in another way. Using pigeons as subjects, a discrete trial procedure, and key pecking as the criterion response, Jenkins demonstrated that the positive stimulus elicits twice as many responses to extinction following discrimination training as does the same stimulus following an equal number of continuous reinforcements. On the basis of an ingenious arrangement of controls, Jenkins is able to conclude that the increased resistance to extinction is not due to the fact that the positive stimulus is eventually discriminated from the negative, but rather to the occurrence of nonreinforced responses in the process of learning the discrimination.

FRUSTRATIVE NONREWARD THEORY AND PREDISCRIMINATION EFFECTS

In the Ward experiment, subjects were exposed before a discrimination

TABLE 1

MEAN DAY ON WHICH CRITERION OF DISCRIMINATION REACHED BY GROUPS FOLLOWING 0, 12, OR 48 PREDISCRIMINATION B+ W+ TRIALS

Condition	Group			*F*
	0	12	48	
Running time				
Easy criterion	8.0	6.5	4.5	3.64*
Hard criterion	12.8	10.3	8.6	5.26**
Goal entry time				
Easy criterion	7.2	4.3	4.1	3.22*
Hard criterion	10.9	8.1	6.1	6.14**

* $p < .05$.
** $p < .01$.

reasoning for the intermediate case is that both r_R and r_F are being evoked by B at the start of discrimination, but s_R elicits approach and s_F elicits avoidance. When, after 32 trials, we switch from prediscrimination B± to the B+W− discrimination, B+ evokes both approach and avoidance tendencies through s_R and s_F, and W−, new in the discrimination, evokes mainly generalized approach (r_R) and little avoidance (r_F). When the shift is from B± to B−W+, there is a strong tendency for s_F to elicit avoidance in relation to the now negative B− (generalizing very little to W+) and a strong tendency for s_R to elicit approach to the new, now positive, W+. This latter situation is much more favorable for discrimination. The prediction: After an *intermediate* number of partial rewards of an approach response to B, B+W− will produce *slower* discrimination than will B−W+.

After a few B± trials, enough to produce some r_G but little or no FE and no r_F, there should be no difference in the rate of the B+W− and B−W+ discriminations.

Continuous Reward of a Response to One Discriminandum Prior to Discrimination

We deal here with the underlined conditions on the right of Figure 4: continuous (B+) prediscrimination experience followed by a discrimination in which B remains positive (B+W−) or becomes negative (B−W+). The expectations here are most interesting when we compare them to the predictions just made for prediscrimination partial reward. Briefly, with increasing numbers of rewards of approach responses to the prediscrimination stimulus, B+ should evoke increasing r_R, and nonreward in the subsequent discrimination should be increasingly frustrating. Since r_R generalizes strongly from B to W, both B and W should evoke r_R quite strongly and almost equally in the discrimination. Nevertheless, the FE should be somewhat greater when B is made the negative stimulus in the discrimination than when it stays positive. However, in either case (B+→B+W− or B+ →B−W+), the flat gradient of generalization of positive tendencies leads to a prediction which is the same as the earlier Ward result: increasing the number of continuous prediscrimination rewards should produce faster discrimination.

In contrast to the center panel of Figure 6 in which the relationships at the middle and end points are reversed, the differences between the two curves in the top panel are shown as consistently in favor of faster discrimination for B+→B+W−. This is not a prediction based on frustrative nonreward factors alone. In fact, part of this is not a prediction at all. There is evidence in some unpublished data[5] that a switch from B+ in prediscrimination to B− in discrimination does produce a greater FE in Runway 2 than a switch from B+ to W−, as expected (Amsel & Hancock, 1957); however, an immediate generalization decrement of the Runway 1 response in switching from B to W weakens response to W, which speeds the apparent separation of the B+W− discrimination curves and retards the separation of the B−W+ curves. The suggestion is that, in the case of prediscrimination continuous reinforcement, the factor of generalization decrement, operating against the somewhat greater FE when the prediscrimination stimulus becomes negative in the discrimination, should maintain the relative advantage of B+→B+W− over B+→B−W+

---

[5] The reference is to data collected by Susan Hicks at Newcomb College, Tulane University, as part of an undergraduate honors project.

for all numbers of prediscrimination trials. In prediscrimination partial reinforcement (middle panel of Figure 6), since there is a switch-over in relative positions from 32 to 120 trials, the factor of generalization decrement should reduce the apparent difference at the intermediate point and increase the reversed difference at the 120-trial point.

While some of the implications of the analysis of variables represented by the schema in Figure 4 are currently being tested using a runway apparatus with amplitude changes as the response measures, we plan also to test these ideas in terms of rate of discrimination in two-choice situations with single presentation of discriminanda. In these latter situations, generalization decrement effects due to introduction of new stimuli would still be expected to effect running speed; however, they might have less effect on a correct-choice indicant of speed of discrimination, and provide an additional check on the implications of this position.

We have been looking at situations in which the response to a single stimulus cue is partially or continuously rewarded some number of times, and this stimulus then becomes either the positive or the negative cue in a discrimination. The last two sections of this paper will outline some of the factors operating in prediscrimination experience involving both discriminanda.

Partial versus Continuous Reward of Responses to Both Discriminanda Presented Separately Prior to Discrimination

The bottom panel of Figure 6 summarizes predictions about the rate of discrimination learning following prediscrimination experience with *both* discriminanda presented separately. I am suggesting that B+W+→B+W− operates essentially like B+→B+W− (or B+→B−W+), except that it would not be susceptible to generalization effects, since neither B nor W is new in the discrimination. There is some evidence related to this condition from the Shoemaker and Ward experiments, but only for the part of the "function" covering the first two points. Analysis of B±W± as a prediscrimination condition indicates a decreasing monotonic relationship between rate of discrimination and number of such prediscrimination trials. The reasoning is simply that partial reinforcement of approach responses to both stimuli will result in conflicting approach and avoidance response tendencies to *both* B and W at the 32-trial stage; and the elicitation of approach responses to s_F in connection with both B and W at the 120-trial stage. While the former represents a difficult base on which to build discrimination, as compared with B+W+, the latter should be even worse for discrimination, since it involves the mechanism for the PRE (s_F connected to approach) elicited by the negative stimulus in the discrimination. Consequently, a prolonged period of continued approach to both B+ and W− would be expected in the discrimination phase.

Discrimination Reversal as a "Prediscrimination" Phenomenon

The evidence on discrimination reversal is certainly not unequivocal; however, the weight of it suggests a U shaped relationship between ease of discrimination learning (B+W−) and number of prior B−W+ trials. This would look like the B±→B+W− prediction (middle panel Figure 6) in direction. Reversal, at least up to the 32-trial point, would produce successively slower learning; but reversal after many postcriterion trials, e.g., the 120-trial point, would facilito both of the stimuli, and approach responses in the presence of both were always rewarded. The Ward experiment represents continuous (prediscrimination) reward experience with both discriminanda. Now let us extend frustrative nonreward theory to a class of prediscrimination treatments broader than that represented by the Ward experiment.

Consider how discrimination learning will be affected not only by *continuous* prediscrimination reward experience but also by *partial* prediscrimination reward experience in relation to *one* or to *both* of a pair of discriminanda. I will restrict the discussion, in this paper, to a black-white, successive discrimination (single presentation of stimuli). The problem is to make predictions about the course of such a discrimination on the basis of some history of consistent or of partial reward experience in relation to one or both of the eventual discriminanda. Discrimination is measured in terms of amplitude (running time) changes to the positive and negative discriminanda.

Table 2 lists several possible kinds of exposure to discriminanda prior to a black-white discrimination. As we examine these cases, it is clear that theorists of learning have been interested mainly in prediscrimination experience involving both discriminanda. An early example is the continuity versus noncontinuity experiments which varied amount of presolution experience with the discrimination problem itself (B+W−) or with its reverse (B−W+), and with receptor orienting activity in relation to the discriminanda (e.g., Ehrenfreund, 1948; Gatling, 1951; Krechevsky, 1932, 1938; Lashley, 1942; Spence, 1940, 1945). More recent work on prediscrimination treatments as they influence later discrimination learning has taken several forms including Harlow's (e.g., 1949, 1950, 1957) experiments on learning sets, Lawrence's (1949, 1950, 1952) studies on acquired distinctiveness of cues, and the work relating overlearning to discrimination reversal and transfer (e.g., Birch, Ison, & Sperling, 1960; Capaldi & Stevenson, 1957; Pubols, 1956; Reid, 1953). These newer interests in prediscrimination experience have been somewhat less oriented than the earlier ones toward theoretical argument. In fact, they have to some extent tended to bring together the earlier wide split between S-R and cognitive approaches to discrimination, while investigating mediational processes affecting discrimination. Perhaps a point has been reached where it may be profitable to begin looking at the prediscrimination-discrimination sequence in a way which will bring the cognitive-expectancy position even more into the fold of an S-R (conditioning-expectancy) interpretation. It seems to me that a treatment of discrimination phenomena in terms of anticipatory (conditioned) mechanisms does just this.

Consider the following "attributes" of prediscrimination exposure to a single stimulus: (*a*) The prediscrimination stimulus may become either the positive or the negative stimulus in the eventual discrimination. (*b*) The re-

TABLE 2

KINDS OF PREDISCRIMINATION EXPERIENCE IN RELATION TO A BLACK (+) WHITE (−) DISCRIMINATION

One Prediscrimination Stimulus	
B+	Continuous Reward, Positive Stimulus
W+	Continuous Reward, Negative Stimulus
B±	Partial Reward, Positive Stimulus
W±	Partial Reward, Negative Stimulus
Two prediscrimination stimuli	
B+ W+	Continuous Reward, Both Stimuli
B± W±	Partial Reward, Both Stimuli
B− W+	Discrimination Reversal

sponse to this prediscrimination stimulus may be partially or continuously rewarded. (*c*) The number of prediscrimination exposures to the stimulus-response-reward (or nonreward) sequence may be large, small, or of some intermediate value.[4] Figure 4 shows the various possible combinations of *a*, *b*, and *c* for a black-white discrimination. The 12 cells of this schema represent a matrix of possible experimental conditions, each counterbalanced for absolute effects of color.

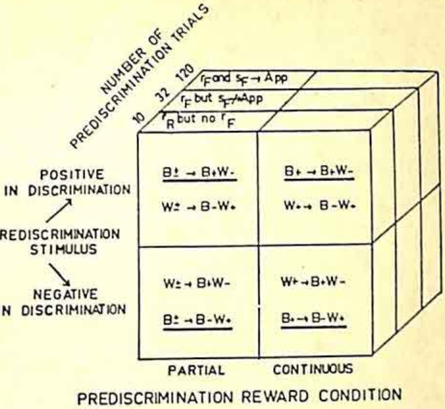


FIG. 4. Schema showing 24 prediscrimination-discrimination sequences (12 conditions, each counterbalanced) resulting from three attributes of prediscrimination exposure to a single discriminandum.

What can be predicted from a behavior theory with frustrative nonreward premises about the effects of prediscrimination experiences in relation to a single discriminandum on subsequent black-white discrimination? Is it possible to show that all combinations of *a*, *b*, and *c* and also effects of prediscrimination experience involving both discriminanda will be deducible from a relatively small number of principles? In the next four sections, I will look at four kinds of prediscrimination treatment and, for each one, try to see what the theory leads us to expect.

Partial Reward of a Response to One Discriminandum Prior to Discrimination

Our guiding conception, when the prediscrimination treatment involves partial reinforcement and a sufficiently large number of trials, is that anticipatory frustration produced stimuli (s_F) become associated with approach responses. (Miller, 1960, has suggested a similar mechanism—"learning to advance to the cue of fear"—to account for the learning of resistance to fear.) The reasoning is that while s_F will at first elicit responses antagonistic to approach, it will finally become connected to approach tendencies. This means that at some stage in partial reinforcement s_F will become part of a stimulus complex evoking approach, and earlier evidence of conflict will disappear.

---

[4] The values for number of trials, 10, 32, and 120, which appear in Figures 4, 5, and 6, and in the text in various places, are specified here for purposes of exposition and because we are now conducting experiments which involve these numbers of prediscrimination trials. However, these numbers of trials only represent guesses, at the present time, as to when the various stages of practice are reached. Rather than be pinned down to these numbers of trials in the predictions for discrimination learning which follow, I would prefer to set up independent indices of the existence of one stage or another and to predict discrimination outcome on the basis of these indices rather than some number of trials. The index for the first stage is relatively safe and simple. It is a stage at which nonreward produces no evidence of frustration as measured by the FE in a double runway, or in any other way. If there is no frustration, there can be no r_F. The presence of r_F takes us into the second stage. The choice of a number of trials to designate the second stage is most difficult since it must represent a very critical phase in partial reinforcement: enough trials for r_F to be present but not so many that s_F is associated with approach. The independent indicator of the middle stage (r_F, but s_F not associated with approach) is increased variability of behavior soon after nonreward produces the FE (Amsel, 1958). Stopping partial reward acquisition at such a point defines this middle stage. The late stage (identified here, conservatively, as 120 trials) has a clear independent definition. It exists (*a*) when variability decreases again, but more particularly (*b*) when partial reward performance exceeds continuous reward performance in vigor (see Goodrich, 1959; Haggard, 1959; Spence, 1960; Wagner, 1961a). While it might appear from earlier partial reinforcement experiments that 32 trials of partial reinforcement are sufficient to get us to the last stage and produce the PRE (and therefore are too many trials for our intermediate condition) experiments demonstrating PRE with this many acquisition trails (e.g., Sheffield, 1949; Weinstock, 1954) have usually involved an initial period of about 10 continuous rewards. Our procedure involves partial reward from the outset.

Of the four stages of practice in partial reinforcement previously outlined, three are theoretically differentiable with respect to the involvement of r_F, and therefore s_F: a stage where r_R has been conditioned, but not yet r_F (10 trials); a stage where both r_R and r_F have been conditioned and their stimuli, s_R and s_F, still evoke competing (approach and avoidance) response tendencies (32 trials); and a stage where both r_R and r_F are conditioned and the stimuli they produce, s_R and s_F, both evoke approach tendencies (120 trials). This conceptualization of three stages in partial reinforcement in regard to r_F—s_F, along with an assumption borrowed from Miller's analysis of conflict and displacement (1944, 1948), will provide the basis for some deductions which I find interesting.

The assumption from Miller's analysis of conflict, for which there is good experimental support, is that the generalization gradient for positive (approach) tendencies is flatter than that for negative (avoidance) tendencies. If we now identify $s_R \rightarrow$App (which results from r_R) and $s_F \rightarrow$Av (which results from r_F) as such positive and negative tendencies, the three panels of Figure 5, from top to bottom respectively, represent the state of affairs at the beginning of a discrimination, when the discrimination has been preceded by a few, an intermediate number, or a large number of partial reinforcements of a response to the prediscrimination stimulus (B). Represented on the left-hand side of the graph are the relative strengths of approach and avoidance tendencies to the prediscrimination stimulus (B) through the mediating response produced stimuli, s_R and s_F. Shown on the right are the generalized strengths of these tendencies to the stimulus (W) which *will be new* in the discrimination. (If the prediscrimination stimulus were W, then B would be the new stimulus in the discrimination and the B and W positions on the baseline would of course be reversed.)

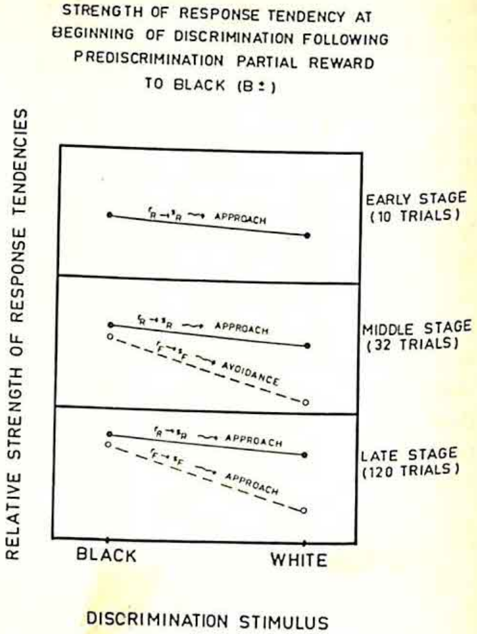


FIG. 5. An application of an assumption from Miller's analysis of conflict to predictions of ease of discrimination learning following various prediscrimination exposures to a single discriminandum.

A simplifying assumption, represented in the bottom panel of Figure 5, is that the strength of approach, elicited through r_F—s_F by B and W in the late stage, is related to the strength of avoidance tendencies elicited through r_F—s_F in the middle stage. I am suggesting that the avoidance tendency to s_F is "neutralized" and has the effect of an approach tendency at this stage.

Some predictions can now be made relative to the conditions in Figure 4, and these will be restricted to the underlined conditions. (The other conditions simply counterbalance for any absolute effects of color and I assume they will not be different in direction.) Again, let me emphasize that the predictions will differ depending on the depth dimension: Was prediscrimination training discontinued at the stage of development of r_R but not r_F, indicated by the point of earliest appearance of FE? Or, was training carried on to the intermediate conflict stage at which s_F evokes avoidance? Or, were a very large number of trials run, allowing s_F to become connected to approach? Outlined below, for each of these cases, are predictions as to which discrimination will be learned faster: that in which the prediscrimination stimulus remains positive in the discrimination (B±→B+W−), or that in which it becomes negative (B±→B−W+).

Let us begin with the case of discrimination following 120 partial rewards. The reasoning (see Figure 5): At the start of discrimination, B will elicit r_R and r_F strongly and about equally. On the other hand, W, which is new in the discrimination, will elicit r_R relatively strongly because of the flat positive gradient, but will elicit r_F very weakly because of the steep negative gradient. This means that when B is made negative in the discriminination after a great many prediscrimination trials, B− will elicit strong r_F—s_F, and s_F will elicit continued approach. This will slow discrimination relative to the condition where B is made positive in the discrimination. Here the W− stimulus which is new in the discrimination elicits very weak r_F. Consequently, there is little s_F to elicit approach, little to counteract the build-up of avoidance to W−, and discrimination should develop more quickly. The theory predicts the following: After a very *large* number of prediscrimination partial rewards of an approach response to B, B+W− will produce *faster* discrimination than will B−W+. This prediction and its relationship to other predictions are shown in Figure 6.

The most interesting and critical comparison to be made for purposes of our analysis is between the 120-prediscrimination-trial case and the intermediate, 32-prediscrimination-trial case. In the 120-trial case s_F is already connected to approach, while in the 32-trial case s_F still elicits avoidance. Consequently the predictions in these two cases are reversed. The

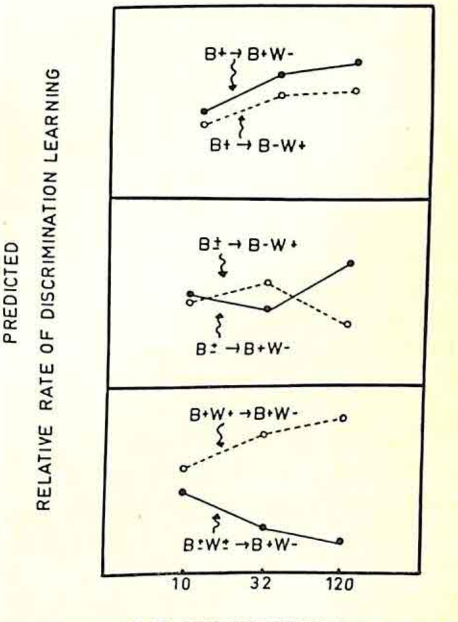


FIG. 6. Predicted rate of discrimination learning following various prediscrimination exposures to one or both of the eventual discriminanda.

tate the formation of the reversed discrimination.

The discrimination reversal studies cited earlier have been conducted in situations involving joint presentation of stimuli. They involve two basic findings. The first concerns ease of discrimination following the first reversal; the second concerns the "learning to learn" phenomenon—that successive reversals shorten the numbers of trials required to criterion. For the first of these findings, there is scant evidence in situations involving single presentation of stimuli (the present case); however the recent Birch, Ison, and Sperling study (1960) is important and relevant.

The conditions of the Birch, Ison, and Sperling experiment are closest to meeting the boundary conditions of a frustrative nonreward analysis of reversal as a "prediscrimination" phenomenon. Their subjects, rats, were given 11 trials each day, the first always positive, the next 10 divided between positive and negative according to random orders. Their measure was response speed rather than choice, and their apparatus was essentially a runway. They were trained initially to approach a white platform and to avoid a black one and, when they reached a criterion of discrimination defined in terms of approach speed, half of the subjects were switched to a B+W− discrimination immediately, the other half after 12 additional days of overtraining to B−W+. The average number of trials to reach the original criterion for both groups was about 50 (well beyond our 32-trial point). This means that the overtrained group had about 180 trials to B−W+ (well beyond our 120-trial point) before it was reversed and learned B+W−. The overtrained group learned the reversed discrimination in significantly fewer trials.

Interpreting their results, Birch, et al. (1960) state: ". . . the difference between the two groups appears to be located in their differential rates of extinction to the previously reinforced cue. No difference in rate of acquisition to the newly positive cue was found in the present data" (p. 39). These results have a rather obvious interpretation in terms of a frustrative nonreward position; but the situation is complex and there are theoretical difficulties to overcome. The obvious part of the interpretation is that the overlearned group has developed stronger r_R to W+, which makes subsequent W− more frustrating, which facilitates the conditioning of W to r_F, which, finally, ensures the evocation of avoidance responses by s_F. This part is fine for the theory; however, there is the matter of what happens to s_F, which presumably develops in the initial discrimination of the Birch, et al. experiment to the negative stimulus (B−) and which is connected to avoidance. Shouldn't the overtrained group have a stronger $s_F \rightarrow$ avoidance tendency in relation to B than the criterion group? And shouldn't this retard the learning of the reversed discrimination in the overtrained group when B subsequently becomes the positive stimulus? This difficulty remains for the theory unless one assumes that r_R (and $s_R \rightarrow$ approach) builds in strength very slowly over many trials (at least over 50) while r_F (and $s_F \rightarrow$avoidance) develops quickly and is maximal by about 50 trials—this being the case despite the fact that the growth of r_F depends upon the prior growth of r_R. This is certainly an interesting possibility.

SUMMARY

This paper begins with an account of recent evidence for a concept of frustrative nonreward in behavior

theory, and presents support for the position that an "active" role must be assigned to nonreward in both motivation and inhibition. This evidence is not only from experiments specifically designed to study frustrative properties of nonreward, but also from a variety of other kinds of experiments, including, among others, studies of extinction as a function of magnitude of reward and number of rewards, and experiments in discrimination learning which demonstrate the relative importance of nonrewarded and rewarded trials.

The importance of an active conception of nonreward in current thinking is discussed with reference to similarities and differences between recent cognitive-expectancy and neo-Hullian interpretations of the partial reinforcement effect. The suggestion is that current S-R conditioning interpretations of these partial reinforcement phenomena, stressing the role of both anticipatory reward and anticipatory frustration mechanisms, and involving a more specific and better worked out language, provide a more workable "expectancy" theory than do cognitive interpretations.

The contention is that current neo-Hullian approaches, employing conditioning-expectancy concepts, go beyond cognitive-expectancy approaches in both specificity and predictive power. As evidence for the merit of such a position, data are presented from discrimination learning experiments which test predictions from a conditioning analysis of nonreward effects, and an extension of frustrative nonreward theory is presented which leads to predictions about the course of discrimination learning following various prediscrimination experiences. These predictions depend on the interaction of three attributes of prediscrimination experience: (*a*) whether prediscrimination exposure is to the stimulus which is eventually positive or negative in the discrimination (or to both); (*b*) whether the prediscrimination treatment involves continuous or partial reward; and (*c*) the number of prediscrimination exposures (trials), representing different stages of development of anticipatory reward and anticipatory frustration.

REFERENCES

ADELMAN, H. M., & MAATSCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, 1955, **50**, 61–65.

ADELMAN, H. M., & ROSENBAUM, G. Extinction of instrumental behavior as a function of frustration at various distances from the goal. *J. exp. Psychol.*, 1954, **47**, 429–432.

AMSEL, A. A three-factor theory of inhibition: An addition to Hull's two-factor theory. Paper read at Southern Society for Philosophy and Psychology, Roanoke, 1951.

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A., ERNHART, C. B., & GALBRECHT, C. R. Magnitude of frustration effect and strength of antedating goal factors. *Psychol. Rep.*, 1961, **8**, 183–186.

AMSEL, A., & HANCOCK, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126-131.

AMSEL, A., & PENICK, E. The influence of early experience on the frustration effect. *J. exp. Psychol.*, 1962, **63**, 167–176.

AMSEL, A., & PROUTY, D. L. Frustrative factors in selective learning with reward and nonreward as discriminanda. *J. exp. Psychol.*, 1959, **57**, 224–230.

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

AMSEL, A., & WARD, J. S. Motivational properties of frustration: II. Frustration drive stimulus and frustration reduction in selective learning. *J. exp. Psychol.*, 1954, **48**, 37–47.

ARMUS, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

Berry, R. N., Verplanck, W. S., & Graham, C. H. The reversal of discrimination in a simple running habit. *J. exp. Psychol.*, 1943, **32**, 325–334.

Birch, D. Discrimination learning as a function of the ratio of nonreinforced to reinforced trials. *J. comp. physiol. Psychol.*, 1955, **66**, 456–464.

Birch, D., Ison, J. R., & Sperling, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

Bitterman, M. E., Fedderson, W. E., & Tyler, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456-464.

Bower, G. H. A contrast effect in differential conditioning. *J. exp. Psychol.*, 1961, **62**, 196–199.

Bower, G. H., & Stocks, D. Graded frustration effect as a function of graded reduction in reward. Paper read at Psychonomic Society, Chicago, 1960.

Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.

Bush, R. R., & Mosteller, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.

Cantor, G. N., & Spiker, C. C. Effects of nonreinforced trials on discrimination learning in preschool children. *J. exp. Psychol.*, 1954, **47**, 256–258.

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

D'Amato, M. R., & Jagoda, H. Effects of extinction trials on discrimination reversal. *J. exp. Psychol.*, 1960, **59**, 254–260.

Denny, M. R., & Dunham, M. D. The effect of differential nonreinforcement of the incorrect response on the learning of the correct response in the simple T-maze. *J. exp. Psychol.*, 1951, **41**, 382–389.

Ehrenfreund, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol. Psychol.*, 1948, **41**, 408–422.

Eninger, M. U. The role of generalized approach and avoidance tendencies in brightness discrimination. *J. comp. physiol. Psychol.*, 1953, **46**, 398-402.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of science.* Vol. 2. New York: McGraw-Hill, 1959.

Ferster, C. B. Withdrawal of positive reinforcement as punishment. *Science*, 1957, **126**, 509.

Ferster, C. B. Control of behavior in chimpanzees and pigeons by time out from positive reinforcement. *Psychol. Monogr.*, 1958, **72** (8, Whole No. 461).

Festinger, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

Festinger, L. The psychological effects of insufficient rewards. *Amer. Psychologist*, 1961, **16**, 1–11.

Fitzwater, M. E. The relative effect of reinforcement and nonreinforcement in establishing a form discrimination. *J. comp. physiol. Psychol.*, 1952, **45**, 476–481.

Gatling, F. A study of the continuity of the learning process as measured by habit reversal in the rat. *J. comp. physiol. Psychol.*, 1951, **44**, 78–83.

Goodrich, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

Grice, G. R., & Goldman, H. M. Generalized extinction and secondary reinforcement in visual discrimination learning with delayed reward. *J. exp. Psychol.*, 1955, **50**, 197–200.

Grove, G. R., & Eninger, M. U. The relative importance of approach and avoidance tendencies in brightness discrimination learning. Paper read at Midwestern Psychological Association, Chicago, 1952.

Haggard, D. F. Acquisition of a simple running response as a function of partial and continuous schedules of reinforcement. *Psychol. Rec.*, 1959, **9**, 11–18.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

Harlow, H. F. Analysis of discrimination learning by monkeys. *J. exp. Psychol.*, 1950, **40**, 26-39.

Harlow, H. F., & Hicks, L. H. Discrimination learning theory: Uniprocess vs. duoprocess. *Psychol. Rev.*, 1957, **64**, 104–109.

Holder, W. B., Marx, M. H., Holder, E. E., & Collier, G. Response strength as a function of delay of reward in a runway. *J. exp. Psychol.*, 1957, **53**, 316–323.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

HULL, C. L., LIVINGSTON, J. R., ROUSE, R. O., & BARKER, A. N. True, sham, and esophageal feeding as reinforcements. *J. comp. physiol. Psychol.*, 1951, **44**, 236–245.

HULSE, S. H., JR. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

HUMPHREYS, L. G. The effect of random alternation of reinforcements on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

HUMPHREYS, L. G. Extinction of conditioned psycho-galvanic response following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–76.

JENKINS, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, **61**, 111–121. (a)

JENKINS, H. M. Resistance to extinction when partial reinforcement is followed by various amounts of regular reinforcement. Paper read at Psychonomic Society, New York, 1961. (b)

KENDLER, H. H., PLISKOFF, S. S., D'AMATO, M. R., & KATZ, S. Nonreinforcements versus reinforcements as variables in the partial reinforcement effect. *J. exp. Psychol.*, 1957, **53**, 269–276.

KRECHEVSKY, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, **39**, 516-532.

KRECHEVSKY, I. A study of the continuity of the problem-solving process. *Psychol. Rev.*, 1938, **45**, 107–133.

LACHMAN, R. The influence of thirst and schedules of reinforcement-nonreinforcement ratios upon brightness discrimination. *J. exp. Psychol.*, 1961, **62**, 80–87.

LAMBERT, W. W., & SOLOMON, R. L. Extinction of a running response as a function of distance of block point from the goal. *J. comp. physiol. Psychol.*, 1952, **45**, 269–279.

LASHLEY, K. S. An examination of the "continuity theory" as applied to discrimination learning. *J. gen. Psychol.*, 1942, **26**, 241–265.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.

LAWRENCE, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.

LAWRENCE, D. H. Learning. *Annu. Rev. Psychol.*, 1958, **9**, 157–188.

LAWSON, R., & MARX, M. H. Frustration: Theory and experiment. *Genet. psychol. Monogr.*, 1958, **57**, 393-464.

LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

LONGSTRETH, L. E. The relationship between expectations and frustrations in children. *Child Develpm.*, 1960, **31**, 667–671.

MARZOCCO, F. M. Frustration effect as a function of drive level, habit strength and distribution of trials during extinction. Unpublished doctoral dissertation, State University of Iowa, 1951.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MILLER, N. E. Learning resistance to pain and fear: Effects of overlearning, exposure, and rewarded exposure in context. *J. exp. Psychol.*, 1960, **60**, 137–145.

MILLER, N. E., & STEVENSON, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.

MOWRER, O. H. *Learning theory and behavior.* New York: Wiley, 1960.

NISSEN, H. W. Description of learned responses in discrimination behavior. *Psychol. Rev.*, 1950, **57**, 121–131.

NORTH, A. J., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

OLDS, J. The influence of practice on the strength of secondary approach drives. *J. exp. Psychol.*, 1953, **46**, 232-236.

OLDS, J. *The growth and structure of motives.* Glencoe, Ill.: Free Press, 1956.

PAVLIK, W. B. The effects of deprivation schedule and reward magnitude on performance during acquisition and extinction. Paper read at Eastern Psychological Association, Philadelphia, April 1961.

PENNEY, R. K. The effects of non-reinforcement on response strength as a function of number of previous reinforcements. *Canad. J. Psychol.*, 1960, **14**, 204–215.

PUBOLS, B. H., JR. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

REYNOLDS, W. F., RICHTER, M. L., & CARLOCK, J. Resistance to extinction as a function of reward magnitude and number of training trials. Paper read at Psychonomic Society, Chicago, 1960.

REYNOLDS, W. F., & SIEGEL, M. H. Drive reversal effect as a function of magnitude of reward. *Psychol. Rep.*, 1961, **8**, 135–141.

ROHRER, J. H. A motivational state resulting from non-reward. *J. comp. Psychol.*, 1949, **42**, 476–485.

ROSENZWEIG, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

ROUSELL, J. Frustration effects as a function of repeated non-reinforcements and as a function of the consistency of reinforcement prior to introduction of nonreinforcement. Unpublished master's thesis, Tulane University, 1952.

SEWARD, J. P. Experimental evidence for the motivating function of reward. *Psychol. Bull.*, 1951, **58**, 130–149.

SEWARD, J. P., PEREBOOM, A. C., BUTLER, B., & JONES, R. B. The role of prefeeding in apparent frustration effect. *J. exp. Psychol.*, 1957, **54**, 445–450.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

SHEFFIELD, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.

SHOEMAKER, H. A. The relative efficiency of rewarded and non-rewarded training in a black-white discrimination problem. Unpublished doctoral dissertation, University of Colorado, 1953.

SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

SOLOMON, R. L. The externalization of hunger and frustration drive. *J. comp. physiol. Psychol.*, 1956, **49**, 145–148.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

SPENCE, K. W. Analysis of the formation of visual discrimination habits in the chimpanzee. *J. comp. Psychol.*, 1937, **23**, 77–100.

SPENCE, K. W. Continuous versus noncontinuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.

SPENCE, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253–266.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavioral theory and learning.* Englewood Cliffs: Prentice-Hall, 1960.

STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.

THEIOS, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, in press.

TYLER, D. W., MARX, M. H., & COLLIER, G. Frustration stimuli in discrimination. *J. exp. Psychol.*, 1959, **58**, 295–301.

WAGNER, A. R. The treatment of inhibition in the more complex behavioral situation. Paper read at American Psychological Association, 1961. (b)

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *J. exp. Psychol.*, 1959, **57**, 130–136.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242. (a)

WARD, J. S. An experimental investigation of the role of frustrative factors in discrimination learning. Unpublished doctoral dissertation, Tulane University, 1962.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

WILSON, W., WEISS, E. J., & AMSEL, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50**, 51–60.

(Received November 20, 1961)

Psychological Review
1962, Vol. 69, No. 4, 329-343

THE SELECTION OF STRATEGIES IN CUE LEARNING[1]

FRANK RESTLE

Michigan State University

In a cue learning problem (discrimination learning, concept formation, and maze learning) the subject chooses one of two or more responses on each trial. The correctness of the response depends on some aspect of the situation at the time of response.

In this paper it is assumed that subjects have difficulty with cue learning problems to the degree that they tend to use strategies (habits, patterns of response) which conflict with the strategy intended by the experimenter. It is often assumed that the subject must also associate correct responses with cues in the situation. However, in most cue learning experiments the subject is instructed or pretrained to make the desired responses before the cue learning process begins. Furthermore, the experimental situation constrains the subject to make exactly one of the available responses per trial. It seems reasonable to assume that cue learning is not so much a matter of the formation as of the selection of responses. The present theoretical discussion begins with a set of strategies and is concerned with the mechanisms by which the subject might select out those strategies intended, and consistently rewarded, by the experimenter.

The term "strategy" is employed in a sense related to the more common terms, "habit" and "hypothesis," to designate a particular pattern of responses to stimuli. Consider, for example, a rat being trained to jump to the larger of two white circles in a Lashley Jumping Stand. Some possible strategies include; jumping to the left side, jumping to the right side, jumping alternately left and right, jumping always to the side last reinforced, jumping to the side which last was punished, jumping in the direction the rat was placed down on the stand, jumping toward the circle with less (or more) acute curvature, jumping toward the side with more white area, jumping toward the side which reflects more total light, and so forth. Each such pattern of behavior, if it occurs in isolation for a sufficiently long sequence of trials, can be identified by the experimenter. It will be seen that there are ordinarily a large number and variety of possible strategies in a cue learning problem, and that a number of strategies may be perfectly confounded. In the example above, jumping to the circle with less acute curvature will always lead to the same result as jumping to the side which reflects more light, since these will both be jumps to the larger circle.

THREE MODELS OF THE SELECTION OF STRATEGIES

The general idea developed here is that the problem gives rise to a set of strategies. The subject uses these various strategies, which at first are selected at random, and attempts to select strategies which will consistently lead to correct responses. For simplicity the discussion is limited to

---

[1] This research was facilitated by the author's tenure as Faculty Research Fellow of the Social Science Research Council, 1959-61.

problems in which there are at least some strategies which are always correct.

Let H be the set of strategies available to the subject at the beginning of the problem. Suppose that some subset C of these strategies are always correct, another subset W are always wrong, and the remainder, I, are sometimes correct and sometimes wrong. For simplicity of presentation, begin with a problem in which the strategies of I are correct and wrong, on successive trials, with independent probabilities all of which are $\frac{1}{2}$, and restrict attention to the two-choice problem.

Presentation of the theory of this conventional discrimination learning problem will begin with a special case in which one strategy is used on each trial. Then an alternative theory will be proposed, in which the subject uses all strategies at once and attempts to narrow down to the correct one. Third, a model will be proposed in which the subject chooses a random sample of all strategies and attempts to narrow down to the correct strategies within his sample. These three models will be shown to be essentially equivalent, in a theorem which may be called the "indifference to sample size" theorem. Following proof of that theorem, some empirical implications of the theory will be derived and compared with relevant data.

One Strategy at a Time

On the first trial of the experiment, the subject selects a certain single strategy from H at random and makes the indicated response. If the response is correct, the same strategy is used on Trial 2. If it is again correct, it is used again on Trial 3, and so forth. If, on any trial, the response is incorrect, the subject returns his strategy to the set H and then chooses a strategy again at random. It is considered possible that the same strategy may be resampled, so that the sampling process is with replacement. Imagine that an error occurs on Trial 2. Then a new strategy is chosen and this strategy is used on Trial 3. If it is correct it is used again on Trial 4. If the response on Trial 3 is wrong, the subject again chooses an hypothesis at random for Trial 4.

Notice that if the subject chooses a correct strategy, he will make no more errors. The correct strategy leads to correct responses, and when the response is correct the subject does not change strategy. Thus, the learning process is terminated by the choice of a correct strategy.

Since sampling is with replacement, the probabilities of choosing a correct, a wrong, or an irrelevant strategy are constant from trial to trial. Let these three probabilities be c, w, and $i = 1 - c - w$.

At the beginning of training the subject chooses a strategy of one of the three types, which have probabilities c, w, and i. After each error he also chooses a strategy, with probabilities c, w, and i. Thus an error is an event which returns the subject to exactly the same condition (in terms of the probabilities of future events) with which he started—it resets the process and makes it begin over. Such an event is called a "recurrent event" (Feller, 1950). Since our experimental interest centers around errors and correct responses, we can limit attention to errors and thus to the theory of recurrent events.

Imagine that at some Trial n the subject makes an error. There is a certain probability f_1 that the next error occurs at Trial $n + 1$, a probability f_2 that the next error occurs at Trial $n + 2$, and in general a probability distribution f_j that the next error after Trial n occurs at Trial $n+j$. In a model of recurrent events like the present one, the distribution f_j is the same for all Trials n on which an error occurs, and is the same as the distribution of probabilities that the first error occurs on Trial 1, 2, etc.

The random process and all of its properties are specified by the distribution f_j. However, the analysis of the process from the distribution f_j will be postponed until two steps have been completed; first, calculation of the distribution f_j for the one-strategy-at-a-time model, and second, the formulation of two other models, very different in assumptions, which also lead to systems of recurrent events with the same distribution. When these steps are completed, the generality of the f distribution will be apparent, and derivations of properties of the process will be justified.

First, f_1 is the probability of an error immediately following another error. The second error can occur because a wrong strategy was chosen (with probability w) or because an irrelevant strategy was chosen and turned out to be wrong (with probability $\frac{1}{2}i$). These are mutually exclusive events and exhaust the possibilities, so that:

$$f_1 = w + \tfrac{1}{2}i.$$

An error occuring on the second trial after the last one, with a correct response intervening, cannot result from the selection of a wrong strategy for the wrong strategy would not lead to an intervening correct response. If the error does occur on the second trial we know that the strategy chosen was not correct. Hence it must be an irrelevant one which led first to a correct and then to a wrong response. In simple discrimination learning, the probability that an irrelevant strategy would lead to a correct and then a wrong response is $\frac{1}{2}\cdot\frac{1}{2} = (\frac{1}{2})^2$. The probability of choosing an irrelevant strategy which gives this sequence, correct-wrong, is therefore:

$$f_2 = i(\tfrac{1}{2})^2$$

By the same reasoning, any value of f_j, $j > 2$, implies that an irrelevant strategy was chosen and then led to $j-1$ correct responses followed by a wrong response. The probability of a string of exactly $j-1$ correct responses each with probability $\frac{1}{2}$, followed by an error with probability $\frac{1}{2}$, is $(\frac{1}{2})^j$. Hence, for $j \geq 2$:

$$f_j = i(\tfrac{1}{2})^j$$

Summarizing the results we have the distribution:

$$f_1 = w + \tfrac{1}{2}i$$
$$f_j = i(\tfrac{1}{2})^j$$

for all $j \geq 2$. The distribution f_j is not a proper probability function because it does not sum to unity. Notice that:

$$\sum_{j=1}^{\infty} (f_j) = w + i \sum_{j=1}^{\infty} (\tfrac{1}{2})^j = w + i = 1 - c$$

In Feller's terms this means that errors are *uncertain* recurrent events, for with probability c the subject chooses a correct strategy and never makes another error. Some of the probability of the f distribution is located at the (improper) point ∞, and the proportion so located is c. This merely reflects the fact that learning occurs in the present model and the subject can eliminate errors. The fact that f_j is not a probability function does not place any serious difficulties in the way of analysis.

All Strategies at Once

We now consider a second model of a subject working on the same discrimination learning problem discussed above, with exactly the same set of strategies H divided into correct (C), wrong (W), and irrelevant (I) subsets. Since a strategy corresponds to any specific cue in the situation and these cues are numerous and finely divisible, consider that the set of strategies consists of a fairly sizeable (though finite) number of strategies, each of which is as likely to be used, and has as much influence on behavior, as every other.

In the all-strategies-at-once model we imagine that the subject begins the learning process by considering the entire set H of strategies simultaneously. In informal terms, he then attempts to narrow down to the correct strategy by a process of elimination. However, we suppose that the subject has only a limited memory and cannot remember all of the strategies at once, but only those with which he is presently concerned. The main disability of the subject in this model is that he may accidentally eliminate the correct strategies. When this happens he is unable to remember what they were, for he can remember only the ones he is still using. When the correct strategies have been lost the subject has no recourse but to begin over with the whole set of strategies. This general idea is now stated more precisely.

On the first trial the subject considers all strategies and chooses his response. The correct strategies and some of the irrelevant ones lead toward a correct response, the wrong strategies and the remainder of the irrelevant ones lead to an error. We suppose that the probability of a response is equal to the proportion of the strategies which lead to that response, so that, if we let N_c, N_w, and N_i be the numbers of strategies in sets C, W, and I, and if we let N_i^- be the number of irrelevant strategies which lead to an error on the next trial, the probability of an error is:

$$f_1 = (N_w + N_i^-)/(N_w + N_c + N_i).$$

When the subject makes this first response we suppose that he eliminates or sets aside all the strategies he does not use. Suppose, for example, that he chooses a correct response. He at the same time eliminates all the wrong strategies and the irrelevant strategies which would have led him to the opposite response. On the next trial he will have only the correct strategies and the surviving irrelevant strategies left.

To complete the discussion and generalize it somewhat we introduce notation for the set of irrelevant strategies which, following any Trial n, first lead to a wrong response on Trial $n + j$. Let $N_{n,j}$ be the number of irrelevant strategies which lead to correct responses on trials $n + 1$, $n + 2, \ldots, n + j - 1$, and an error on trial $n + j$. Now we can write the probability of an error after an error as:

$$f_1 = (N_w + N_{n,1})/(N_c + N_w + N_i).$$

After one correct response, the only strategies which will lead to an error are those $N_{n,2}$ irrelevent strategies which first lead to an error on the second trial. The total strategies left are N_c correct and $N_i - N_{n,1}$ irrelevant strategies which have not already led toward a wrong response. Thus, the probability of an error following one correct response which follows a given error is:

$$f_2 = (1 - f_1)(N_{n,2}/[N_c + N_i - N_{n,1}]) = \frac{N_c + N_i - N_{n,1}}{N_c + N_w + N_i} \cdot \frac{N_{n,2}}{N_c + N_i - N_{n,1}}$$

Extension of this line of argument shows that in general:

$$f_j = \frac{N_c + N_i - N_{n,1}}{N_c + N_w + N_i} \cdot \frac{N_c + N_i - N_{n,1} - N_{n,2}}{N_c + N_i - N_{n,1}} \cdot \ldots$$
$$\cdot \frac{N_c + N_i - N_{n,1} - \ldots - N_{n,j-2} - N_{n,j-1}}{N_c + N_i - N_{n,1} - \ldots - N_{n,j-3} - N_{n,j-2}} \cdot \frac{N_{n,j}}{N_c + N_i - N_{n,1} - \ldots - N_{n,j-1} - N_{j-1}}$$

In this expression the numerator of each term except the last cancels with the denominator of the following term, so that:

$$f_j = N_{n,j}/(N_c + N_w + N_i)$$

In our ideal experiment, $N_{n,j}$, the number of strategies which first lead to a wrong response on Trial $n + j$ after a given error on Trial n, is independent of n and is given by $N_i(\frac{1}{2})^j$. This statement follows from the idea that irrelevant strategies are correct and wrong with independent probabilities of $\frac{1}{2}$. Actually, $N_i(\frac{1}{2})^j$ is the expected number of irrelevant strategies which first lead to an error on Trial $n + j$, but the derivation to follow does not hinge on the distinction.

If we substitute $N_i(\frac{1}{2})^j$ for $N_{n,j}$ in the expression for f_j and simplify, letting i be the proportion of irrelevant strategies in H, we obtain:

$$f_j = i(\tfrac{1}{2})^j.$$

A similar substitution in the expression for f_1 gives:

$$f_1 = w + \tfrac{1}{2}i.$$

These two results are in agreement with the one-strategy-at-a-time model. We have shown that the one-strategy-at-a-time and the all-strategies-at-once models are both systems in which errors are recurrent events, and we have shown that f_j is the same in both systems. Hence, so far as errors are concerned, the two models are identical.

One remark is in order. In the one-strategy-at-a-time model, the probability of a correct response on any trial is 1, $\frac{1}{2}$, or 0 depending on whether a correct, irrelevant, or wrong strategy is being used. The probability of a correct response in the all-strategies-at-once model can take a variety of values; if there are a great many strategies, the possible values of this probability are numerous and closely packed. The two models are not at all alike in the probabilities involved, and the theorem that they give indistinguishable sequences of errors is by no means obvious on intuitive grounds, even though the proof is easy.

The correspondence of the models extends to the interpretation of the parameters. In the one-strategy-at-a-time model the parameters c, w, and i are the probabilities of selecting a correct, wrong, or irrelevant strategy. In the all-strategies-at-once model the parameters c, w, and i are the proportions of correct, wrong, and irrelevant strategies, where elementary strategies are taken to have equal weight. In terms of experimental realizations of the models, these two sets of parameters are not distinguishable.

It is still possible, of course, to tell which model is more reasonable by any technique which goes beyond the choice data. One might, for example, ask human subjects which strategies they are using, or one might use memory tests to see what parts of the situation have been employed. The equivalence of the models is asserted only for the acquisition of a single problem, and only with respect to the overt choices made: the writer will suggest, in subsequent work, that recognition and transfer tests may permit differential predictions.

A Random Sample of Strategies

The two models discussed above are extremes. One would intuitively imagine that the subject would use, not all strategies and not just one, but some sample of strategies. The third model, called the random-sample-of-strategies model, supposes that on the first trial and after each error the subject draws a random sample of all strategies. So long as he makes no errors the subject discards wrong and irrelevant strategies by the same mechanism as that invoked in the all-strategies-at-once model. If the subject makes an error he takes a new independent sample from H and begins the process over.

The theorem to be proved is, if the sampling process is random in a strict sense to be described below and excludes the possibility that the subject might choose the empty sample, then the random-sample-of-strategies model generates a system of recurrent events which is the same as the other two models given above.

We again consider the set H of strategies, composed of equally potent elementary strategies. Suppose that there are N strategies in H of which N_c are correct, N_w are wrong, and N_i are irrelevant. We shall interpret N_c/N as c, the proportion of correct strategies, N_w/N as w, and N_i/N as i.

In a random sampling model of learning one may assign a fixed probability of being sampled to each element and let the sample size vary randomly, or assign a fixed sample-size, supposing that all such samples are equally probable (Estes, 1959.) To eliminate the possibility of the empty sample it is desirable to deal with a fixed sample size n. The random-sampling specification is that all samples of size n are equally probable.

The total number of samples of size n which can be drawn from a population of N elements is:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Similarly, one can draw $\binom{N_c}{n_c}$ samples of n_c correct strategies from the population of N[illegible]gies. For each of these sampl[illegible] orrect strategies, there are $\binom{N}{}$[illegible]ples of n_w wrong strategies, so that there are $\binom{N_c}{n_c}\binom{N_w}{n_w}$ ways of drawing n_c correct and n_w wrong strategies. For each of these combinations there are $\binom{N_i}{n_i}$ ways of drawing n_i irrelevant strategies. Hence the total number of ways one could obtain n_c correct, n_w wrong, and n_i irrelevant strategies is:

$$\binom{N_c}{n_c}\binom{N_w}{n_w}\binom{N_i}{n_i}$$

where, of course, it is assumed that $n_c + n_w + n_i = n$. Since every sample of size n is assumed to be as likely as any other, the probability of drawing a sample of exactly n_c correct, n_w wrong, and n_i irrelevant strategies is:

$$P(n_c, n_w, n_i) = \frac{\binom{N_c}{n_c}\binom{N_w}{n_w}\binom{N_i}{n_i}}{\binom{N}{n}}$$

which is the three-category generalization of the hypergeometric distribution (Feller, 1950).

It was asserted, in the statement of the model, that once the subject has chosen a sample his performance follows the rules of the all-strategies-at-once model. Hence, after any error and the choice of a sample with n_c, n_w, and n_i strategies, we can construct the f_j distribution. We may write the f distribution conditional on the constitution of the sample as:

$$f(j \mid \text{sample with } n_c, n_w, n_i \text{ strategies}) = \left| \begin{array}{ll} (n_w + \frac{1}{2}n_i)/n & \text{for } j = 1 \\ (\frac{1}{2})^j n_i/n & \text{for } j > 1 \end{array} \right.$$

directly from our results on the all-strategies-at-once model. To obtain the unconditional f_j we multiply each conditional f by the probability of the sample, and sum. This gives us the expectation of the conditional f. For f_1 we have:

$$\begin{aligned} f_1 &= \mathbf{E}[(n_w + \tfrac{1}{2}n_i)/n] \\ &= \mathbf{E}(n_w/n) + \tfrac{1}{2}\mathbf{E}(n_i/n) \end{aligned}$$

However, in the hypergeometric distribution the mean proportion of wrong strategies in samples of size n is just the proportion of wrong strategies in the whole population. That is, $\mathbf{E}(n_w/n) = N_w/N = w$. Similar equations hold for irrelevant and correct strategies. Hence:

$$f_1 = w + \tfrac{1}{2}i$$

and, by the same argument:

$$f_j = (\tfrac{1}{2})^j i$$

for j greater than 1. Thus it is shown that the random-sample-of-strategies model with fixed sample size n is a system of recurrent events (this point is obvious and is not proved above) with the same f distribution found in the one-strategy-at-a-time and the all-strategies-at-once models. At this point one can see that the one-strategy-at-a-time model is a special case of the random-sample-of-strategies model with sample size $n = 1$; and similarly, the all-strategies-at-once model is a special case with sample size $n = N$.

One further generalization is possible. Since the final equations of the random-sample-of-strategies model contain no reference to the sample size n, one would get a system of recurrent events with the given f distribution from any mixture of systems which use different sample sizes. One can withdraw the assumption that the sample size is fixed, permitting the possibility that samples are of all different sizes. The only restrictions are that for any n, all samples of size n are equally likely; and there are no empty samples.

The strategy-sampling model differs from Estes' stimulus-sampling model (Estes, 1959) in an important respect. Let the total number of elements (strategies or, in Estes' case, stimulus elements or patterns) be N and the sample size be n. In Estes' formulations, n/N is analogous to the learning rate θ and controls the rate of learning and all the statistical characteristics of the acquisition data. In the present model, learning is independent of the sample size n/N and depends solely on the composition of the basic set H; namely, the proportions of correct, wrong, and irrelevant strategies in H. Thus while the sampling process invoked in this strategy-sampling theory is one of those used by Estes, its place in the theory is entirely different.

In this section it has been shown that three models of the selection of strategies all lead to the same system of recurrent events. Whether or not this is of interest depends largely on whether the resulting description fits the data of cue learning experiments, a question which is considered in the remainder of the paper.

STATISTICAL PROPERTIES OF THE RAW DATA

The discussion above has been restricted to the distribution f_j, the probability that an error follows the last error by exactly j trials. The f distribution can be estimated directly from raw data, but this is neither a conventional nor a very interesting way of describing the data of cue learning. In this section the data generated by this model are analyzed in several of the ways commonly employed by experimenters. Details are given in Restle (1961).

Learning Curve(s)

Consider three versions of the learning curve. One is the succession of correct and wrong responses by an individual subject. A second is the average learning curve of a group of subjects. The third is a corrected or idealized form of the learning curve computed by adjusting a group learning curve (Vincentizing) or by averaging the data of subjects who are selected after the fact for similarity of over-all performance (Spence, 1956).

According to the present theory, the individual data are composed of a sequence of correct and wrong responses in irregular order, followed by an infinite sequence of correct responses. If $w > 0$, there will be somewhat more errors than correct responses on the resolution trials, and there will be a tendency for errors to follow other errors more often than errors follow correct responses. The probability of an error following an error will be $(w + \frac{1}{2}i)/(w + i)$ whereas the probability of an error following a correct response is $\frac{1}{2}$. An individual subject will produce such below-chance behavior for some block of trials and then abruptly, after some error, will either happen on a correct strategy (in the one-strategy-at-a-time model) or will begin a process of elimination which ends up with all correct strategies. This is an extreme form of the "discontinuous" or insightful learning curve. Unfortunately it is difficult to decide whether any individual subject does or does not exhibit this pattern, so data are usually combined.

The group learning curve is merely the average of a set of (theoretically) discontinuous individual curves. If all subjects have the same parameters c, w, and i, they will nevertheless happen to master the problem at different trials, so that the average learning curve is gradual. Its mathematical form is complex but in general appearance, the group learning from this theory resembles the common "growth" curve (Restle, in 1961).

If one selects, from a larger group of subjects, those who make the same number of total errors, or those who reach criterion at the same time, and averages their learning curves; or if one rescales learning trials as by the Vincent-curve method, the resulting curve will usually be **S** shaped. The writer has investigated this question by generating data which arise directly from the assumptions of the present model, by use of tables of random numbers. When, from a large set of such data, one selects a subgroup of "subjects" who all make the same number of errors, or who reach criterion at about the same trial, and average the performance within such a subgroup, the result is an **S** shaped curve. Exact Vincentizing produces a flat (stationary) learning curve before criterion.

The reason for the **S** shaped curves is not difficult to find; the above methods of selecting or rearranging the data tend to put the (theoretically random) times of solution close together. If the times of solution are put exactly together a step-function should result, but if the times of solution are only grouped close to one another, the step is blurred and an **S** shaped curve results (Spence, 1956).

Summary Statistics of the Data

In the present theory the actual trial on which learning takes place is not in any way fixed, but depends upon the random outcome of the process of selecting strategies. The result is large intrinsic variance in the acquisition phase, which can most conveniently be described in terms of the total errors made by each subject. If each subject is trained until a long sequence of consecutive correct responses has been obtained, one is reasonably sure that the number of errors made approximates the theoretical total errors.

On the first trial or after any error the subject may either make another error, sometime later, or he may make no more errors. The probability that the subject will make at least one more error, the first one in exactly j trials, is f_j. Thus the total probability of at least one more error is:

$$\sum_{j=1}^{\infty} (f_j) = 1 - c$$

With probability c the subject never makes another error at all. From this it is not difficult to show that the probability of exactly k errors is $(1 - c)^k c$. This is the geometric distribution which has mean:

$$\mathbf{E}(k) = (1 - c)/c$$

and variance:

$$\mathrm{Var}(k) = (1 - c)/c^2.$$

The standard deviation of the distribution of error scores should be nearly equal to its mean, according to the theory, and the distribution should show an extreme positive skewness.

Provided that irrelevant strategies are correct just half the time at random, trials-to-criterion behaves very much like total errors.

Methods of Estimating the Parameters w *and* c

An important step in any mathematical development of a learning theory is the estimation of parameters. Fortunately, in the present model quite simple and efficient estimates are available. A maximum-likelihood estimator of c is given by:

$$\hat{c} = 1/(\bar{T} + 1)$$

where $\bar{T}$ is the mean total error score of a group of subjects. The variance of this estimate, with N subjects in the group, is:

$$\mathrm{Var}(\hat{c}) = c^2(1 - c)/N$$

It is also possible to estimate w by the maximum-likelihood method. As was mentioned earlier, a high frequency of consecutive errors in the presolution phase is an indication of a relatively large proportion of wrong strategies, whereas a chance frequency of consecutive errors is an indication that there are relatively few wrong strategies; provided that the irrelevant strategies are correct and wrong strictly at random. The method is to count "Trial 0," an imaginary trial before training begins, as an error. Then for each subject divide this expanded set of errors into M_0 errors which are followed by correct responses and M_1 errors which are followed by errors. Then computing the means of these statistics one has:

$$\hat{w} = \frac{\bar{M}_1 - \bar{M}_0 + 1}{\bar{M}_0 + \bar{M}_1}$$

as a maximum-likelihood estimator of w (Restle, 1961).

SOME VARIATIONS OF THE BASIC MODEL

The theory stated above is so restrictive in its assumptions that it cannot be applied with success to a great many experimental data. For example, with the theory above, one must expect the standard deviation of error scores to be just slightly less than their mean, and this is not observed with any great regularity. Furthermore, animal subjects often show a strong tendency to remain with one strategy even if it is wrong or irrelevant, for a fairly long string of trials. The model given above says that the subject starts over after each error, and this simply is not a plausible assumption for animal studies.[2]

Perseveration

The model is quite easily generalized to take some account of perseveration of strategies. In place of the assumption that the subject always resamples after an error, one supposes that resampling occurs with some probability r, which may be considered a constant for given experimental conditions and also a constant during the presolution period. The subject chooses some strategy or sample of strategies at Trial 1, whence the probability of making zero errors is c. The probability that any actual error is the last one is only rc, the joint probability that (*a*) the subject resamples and (*b*) having resampled, hits on a correct strategy (in the one-strategy-at-a-time model) or goes into a terminal process of narrowing down (in the all-strategies-at-once or the random-sample-of-strategies models.) For this model with perseveration the distribution of total error scores is:

$$P(0) = c$$
$$P(T) = (1 - c)(1 - rc)^{T-1}rc.$$

where $T > 0$. The value of this model in fitting animal data is illustrated by analyzing a set of error scores reported in detail by Harlow, Harlow, Rueping, and Mason (1960), (See Restle, 1960). Harlow's distributions show a relatively high frequency of solutions with zero errors, along with a spread to very large error scores, in original learning by baby monkeys. The result is that the standard deviations of the distributions are somewhat larger than the means. Harlow et al. remarked on the high degree of perseveration shown by some monkeys. Even more striking is Warren's investigation (Warren, 1959b) of the discrimination learning of cats, which show strong perseverative tendencies and which give distributions of error scores with standard deviations much larger than the corresponding means.

It is not at all sure that perseverative effects can be reduced to the single parameter r, and the writer's investigations are entirely insufficient to support the assumption. Within the model there are several symptoms of perseveration; a high frequency of zero error scores relative to the remainder of the distribution, higher standard deviation than mean errors, and consecutive runs of responses which follow a wrong or irrelevant strategy, such as a position habit, object preferance, etc. One can imagine that perseverative tendencies would be strong whenever the incentive for learning is not entirely effective, or when the subjects (perhaps from partial reinforcement) have developed an expectation that rewards cannot always be attained. Something of this last effect, in college students, is suggested by the results of Morin (1955).

---

[2] The writer is indebted to Marvin Levine for pointing out the importance of this factor in animal discrimination learning.

Complexity of Correct Strategies

In the model above it is tacitly assumed that the problem can be solved on the basis of a single correct strategy. This seems a reasonable assumption in the case of simultaneous discrimination. However, when two situations are presented on successive trials it is a question whether complete solution involves learning one strategy (turn right to white and left to black) or two strategies, one for each situation. If two strategies are learned independently and with the same parameters, then the distribution of errors is the distribution of the sum of two random variables (total errors in each situation) each of which has a geometric distribution. This results in a negative binomial distribution (see Restle, 1961) with mean:

$$\mathbf{E}(T) = 2(1 - c)/c$$

and variance:

$$\mathrm{Var}(T) = 2(1 - c)/c^2.$$

These statements follow at once from the fact that the mean of the sum of two independent variables is the sum of the means, and the variance of the sum is the sum of the component variances. With small c the standard deviation will be about $1/\sqrt{2}$ or about 0.7 of the mean.

An interesting fact can be brought forward in favor of this argument, though the writer has not been able to make an exhaustive test. In a case of successive size discrimination by college students (Restle, 1955) the writer found ratios of standard deviation/mean of 1.1 and 0.9 for two problems—both very close to unity, hence in agreement with the original model and with the notion that a single strategy is sufficient to solve the problem. In a comparable study using rats in a successive black-white discrimination, Amsel (1952) obtained data on three groups in which the ratio of standard deviation to mean was 0.65, 0.68, and 0.72. The rat results are in excellent agreement with the hypothesis that the rats had to use two separate strategies to solve the problem. One might guess that rats do not integrate the two habits, right turn to white and left turn to black, into a single cognitive structure, whereas humans might make just such an integration. The present analysis may be useful in throwing further light on the nature of successive discriminations for various species of subjects.

ADDITIVITY OF CUES

In several papers regarding another theory the writer has discussed the additivity of cues (Restle, 1955, 1957, 1958, 1959a; see also Bourne & Restle, 1959, and Trabasso, 1960). In simple terms the experiment involves three groups; one learns a problem based on a set A of cues, the second learns a problem based on the set B of cues, and the third learns a problem which can be solved using either A or B cues disjunctively. If the sets A and B are separate then the third set $A \bigcup B$ should have measure $m(A \bigcup B) = m(\mathrm{A}) + m(\mathrm{B})$, (see Trabasso, 1960). Experimental results have been reconciled with an S-R theory involving adaptation of irrelevant cues (Restle, 1955).

Generally speaking, the calculations from the present model are in good numerical agreement with those from the "adaptation" model. Reanalysis of three of the most satisfactory sets of data used before, and one set not previously used, are reported here. The other data previously discussed are so fragmentary that analysis will hardly be fruitful; and Trabasso (1960) used such difficult problems that many of his subjects failed to learn at all on the more difficult problems, making analysis by the present model unsatisfactory. In general, the estimates reported below are not the maximum-likelihood estimates because the subjects were not run to a strong criterion. However, since learning was nearly complete, the approximate estimates are adequate.

Scharlock (1955; see also Restle, 1957) ran a place-versus-response experiment with rats in which the relative weight of correct place (extra-maze) and correct response (intra-maze) strategies can be estimated. He also ran one group with both place and response strategies correct, and a group with response strategies correct and no place cues present. In the calculations below it is assumed that the number or wrong strategies of a given type (place or response) equals the number of correct strategies. This is a symmetry assumption which, while not strictly appropriate for Scharlock's experiment, is needed to permit prediction, since separate estimates of wrong strategies cannot be made on the available data.

Rats learned to go to the same place (using extra-maze cues), with response cues irrelevant, making an average of 9.7 errors. Using the estimate $\hat{c} = 1/(1 + \bar{T})$, one estimates that the correct place strategies make up $1/10.7 = .093$ of the total set of strategies. Other rats learned to make a constant response to different places, averaging 6.7 errors. We estimate that correct response strategies constitute $1/7.7 = .130$ of the total set.

The simplest interpretation of the experiment is that when both place and response strategies would work, the proportion of correct strategies would be the sum of the proportions of place and response strategies, since the same total set of strategies is available. One predicts that with place-plus-response learning, $c = .093 + .130 = .223$. We have the formula that the expected errors to solution is $(1 - c)/c$. Hence the expected mean errors is $.777/.223 = 3.4$; the obtained mean was 4.0, which is adequately close. The discrepancy can partly be explained by the fact that the place-learning and response-learning groups were not run to a high criterion and probably would have made more errors if tested longer. The fast-learning place-plus-response group would likely not have made more errors, since their performance was excellent at the end of the training given.

Another group learned a fixed response in the absence of any good place cues. We assume that this group simply had neither correct nor incorrect place strategies. Its predicted proportion of correct strategies is $.130/1 - 2(.093) = .130/.814 = .160$. The expected errors is $.840/.160 = 5.25$; the observed mean was 5.0. Both predictions are close to the obtained results, well within sampling deviations.

Warren (1959a) has reported data on monkeys in an experiment analogous to Scharlock's study of rats. Warren's monkeys had to learn position habits in some problems and object-discriminations in other problems. Warren also used a response-plus-object problem (object discrimination with the objects left in the same place each trial) and a pure response problem (e.g., choose the left one of two indistinguishable objects.) Predictions followed the same formulas as above, and were extremely accurate: the proportions of object and response strategies were estimated from behavior on object (position-varied) and position (object-varied) problems. When these values were added to predict object-plus-position, the prediction was a mean of 0.67 errors, whereas Warren observed 0.63. For pure response learning, the model predicted 3.25 mean errors and the observed value was 3.04. These agreements between theory and data are well within the range of probable sampling error.

Similar results were obtained by analyzing Warren's experiment on the additivity of color, form, and size cues (Warren, 1953). The analysis is substantially the same as that given in a previous paper (Restle, 1958). Learning data are collected on problems involving only color, only form, or only size differences between the objects. These data are used to calculate the proportional weights of the three sources of strategies, and the resulting values are recombined to compute predicted learning rates for problems involving two or more dimensions; for example, discrimination of a red triangle from a green circle involves color plus form cues. In each calculation account is taken of the greater total number of strategies involved in problems with added cues, and it is assumed that whenever a set of correct strategies is introduced by adding cues, an equally large set of wrong strategies also enter the situation.

Calculations on Warren's (1953) data afforded four predictions of total errors to solution. The four predictions were wrong by −17, 0, 5 and 10 percent respectively, and all of the errors can reasonably be attributed to sampling variations.

The hypothesis of additivity of cues has also been applied to human learning (Restle, 1959a) in an experiment which required subjects to learn differential verbal responses to consonant syllables. The procedure was simple concept formation with individual letters of the syllables used as cues, and the same sort of additivity of cues as above. The adaptation model gave quite accurate predictions and the present model is, if anything, slightly more accurate; it predicts (given data on the two cues separately) that the added-cue group should make an average of 5.26 errors, whereas the observed mean was 5.25. Considering the variability of the data, the extreme closeness is coincidental.

COMPARISON WITH OTHER THEORIES

The three models of the selection of strategies are conceptually similar to the writer's adaptation theory of discrimination learning (Restle, 1955). The "strategies" of the present model resemble the "cues" of the earlier theory. The strategy-selection model has, as a theorem, that the rate of learning depends upon the proportion of correct strategies. A similar idea is expressed in the adaptation model in which it was assumed that the rate of learning (θ) would depend upon the proportion of relevant cues.

Several of the serious faults of the adaptation theory are corrected in the theory of strategy-selection. First, in the adaptation theory the subject was supposed to begin conditioning relevant cues and adapting irrelevant cues, with a rate θ equal to the proportion of relevant cues, right from the first trial of training. Of course, at that first trial the subject has no possible way of knowing which cues will turn out to be relevant—it would be possible for the experimenter to change his mind after the first trial is complete. Hence, the adaptation theory could apply only to a prescient subject. The selection-of-strategies theory does not have the subject treat different strategies differently except on the basis of trials already completed, hence avoids the absurdity. Second, the idea that the structure of the problem (proportion of relevant cues or correct strategies) controls the rate of learning was only a simplifying assumption, with no justification, in the adaptation model; but is an inescapable theorem of strategy-selection. Third, the adaptation theory yielded a determinate learning curve in the sense that $p(n)$ was exactly specified for a given θ and $p(1)$. The variance of typical data was in large part left unexplained by the adaptation model and had to be attributed to individual differences, even though it is notoriously difficult to find any strong predictor of such learning. In the strategy-selection model learning itself is a random event and the model generates variability comparable with that obtained in the data.

Despite these important differences, the present theory is close enough to the adaptation theory to make it possible to carry over many of the theoretical insights, though in modified form. The one type of prediction discussed in this paper was additivity of cues, but other ideas such as the proposed basis of learning sets (Restle, 1958) and various quantitative relationships in concept identification (Bourne & Restle, 1959) can be recast in the mold of the selection of strategies. The many similarities and few differences in predictions must be studied in detail and cannot be discussed here, except to mention that most of the results of the Bourne and Restle paper can be reproduced using the strategy-selection model.

In more general terms, it may be remarked that the strategy-selection model is similar to the ideas of Lashley (1928) and Krechevsky (1932) in general intent. In comparison with stimulus-sampling theoreies (Restle, 1959b) the strategy-selection model is like theories of the observing response (Atkinson, 1959a, 1959b; Wyckoff, 1952), except that in the strategy-selection model there is no conditioning, only observing (selecting). The mathematical structure of the model is very close to that of Bower's (1960) one-element association model.

SUMMARY

A stochastic model for the solution of cue learning problems by the selection of strategies was stated and developed. Errors were shown to constitute a system of uncertain recurrent events in Feller's sense. Three models, one-strategy-at-a-time, all-strategies-at-once, and a-random-sample-of-strategies, were formulated and shown to yield the same system of recurrent events, to be identical in terms of data. This is the independence-of-sample-size theorem. The basic distribution of the recurrent-events system was used to generate a description of the data, and also a method for estimating the proportions of correct and wrong strategies in a problem. Variations of the model, which take account of perseveration and complex strategies, were indicated. Experimental evidence, mainly on the additivity of cues, was discussed.

REFERENCES

AMSEL, A. Rate of learning a visual brightness discrimination as a function of discriminanda durations. *J. comp. physiol. Psychol.*, 1952, 45, 341–346.

ATKINSON, R. C. The observing response in discrimination learning. Technical Report No. 4, 1959, University of California, Los Angeles, Contract Nonr 233(58). (a)

ATKINSON, R. C. A theory of stimulus discrimination learning. Technical Report No. 1, 1959, University of California, Los Angeles, Contract Nonr 233(58). (b)

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, 66, 278–296.

BOWER, G. H. Properties of the one-element model as applied to paired associate learning. Technical Report No. 31, 1960, Stanford University, Contract Nonr 225(17).

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959.

FELLER, W. *An introduction to probability theory and its applications.* (1st ed.) New York: Wiley, 1950.

HARLOW, H. F., HARLOW, M. K., RUEPING, R. R., & MASON, W. A. Performance of infant Rhesus monkeys on discrimination learning, delayed response, and discrimination learning set. *J. comp. physiol. Psychol.*, 1960, 53, 113–121.

KRECHEVSKY, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, 39, 516–532.

LASHLEY, K. S. *Brain mechanisms and behavior.* Chicago: Univer. Chicago Press, 1928.

MORIN, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, 49, 343–351.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, 62, 11–19.

RESTLE, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychol. Rev.*, 1957, 64, 217–228.

RESTLE, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, 65, 77–91.

RESTLE, F. Additivity of cues and transfer in discrimination of consonant clusters. *J. exp. Psychol.*, 1959, 57, 9–14. (a)

RESTLE, F. A survey and classification of learning models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. (b)

RESTLE, F. A note on the "hypothesis" theory of discrimination learning. *Psychol. Rep.*, 1960, 7, 194.

RESTLE, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, 26, 291–306.

SCHARLOCK, D. P. The role of extramaze cues in place and response learning. *J. exp. Psychol.*, 1955, 50, 249–254.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

TRABASSO, T. R. Additivity of cues in discrimination learning of letter patterns. *J. exp. Psychol.*, 1960, 60, 83–88.

WARREN, J. M. Additivity of cues in visual pattern discriminations by monkeys. *J. comp. physiol. Psychol.*, 1953, 46, 484–486.

WARREN, J. M. Solution of object and positional discriminations by monkeys. *J. comp. physiol. Psychol.*, 1959, 52, 92–93. (a)

WARREN, J. M. Stimulus perseveration in discrimination learning by cats. *J. comp. physiol. Psychol.*, 1959, 52, 99–101. (b)

WYCKOFF, L. B. The role of observing responses in discrimination learning. Part 1. *Psychol. Rev.*, 1952, 59, 431–442.

(Received November 14, 1960)

Psychological Review
1962, Vol. 69, No. 4, 344–354

SUBJECTIVE ORGANIZATION IN FREE RECALL OF "UNRELATED" WORDS[1]

ENDEL TULVING

University of Toronto

This paper is concerned with organization as a dependent variable in free recall verbal learning. A method will be described for measuring the extent to which subjects' recall of verbal items presented in different orders on successive trials is structured sequentially. The method assumes no knowledge on the experimenter's part as to the sources of organization, and it is thus applicable to free recall learning of any list of verbal items. The effect of repetition on organization, and the relation of organization to amount of recall will be briefly examined in a simple experiment. The method and the data are discussed in terms of the problem of the role of repetition in free recall verbal learning.

VERBAL LEARNING AND ORGANIZATION

The fact that learning occurs under conditions of practice, often apparently through mere repetition of the material, is empirically well known, but theoretically not yet fully understood. Indeed, the question of why repetition leads to better recall has not been raised too often explicitly, although it has been implicit in much experimental and theoretical work.

The apparent lack of explicit interest in this problem seems to be related to the restricted aspects of subjects' behavior which are studied in verbal learning experiments. In a large majority of studies the dependent variables have been based on the operation of counting single unordered responses, assigned to categories such as "correct," "intralist intrusion," "remote backward association," and the like. Given but a single basic response variable of this kind, there is relatively little that can be done with the question of why repetition is effective, other than studying the effect of different independent variables on the rate of learning. One can be justifiably sceptical, however, of our chances of ever understanding the basic process of verbal learning by simply counting "correct" and various classes of "incorrect" responses under a variety of experimental conditions. Bolles' (1959) recent suggestion that, "If we are to understand human learning, we must determine what our *S*s are really doing" (p. 580), can be regarded as an example of growing realization that a critical reappraisal of response variables in verbal learning is overdue.

An important contribution to our understanding of the basic acquisition process might well be provided by response variables based on ordered classifications of data (Miller & Frick, 1949). Such variables are particularly interesting in the light of G. A. Miller's (1956a, 1956b, 1956c), conception of verbal learning as a direct consequence of the process of recoding or organization. Rehearsal or repetition, according to Miller, 1956b), has "the very important effect of organizing many separate items into a single unit" (p. 43). Repetition does not change the basic storage capacity of memory. Rather, organizing processes accompanying repetition lead to an apparent increase in this capacity by increasing the information load of individual units.

---

[1] This research has been supported by the National Research Council of Canada under Grant Number APA-39.

Miller's unitization hypothesis, if found tenable, has important implications for research strategy and theory in verbal learning. One of the difficulties in evaluating the hypothesis lies in the lack of appropriate measures of organization. Organization as a response variable has been investigated by many psychologists (e.g., Bousfield, 1953; Cofer, 1959; Jenkins & Russell, 1952; Rothkopf & Coke, 1961), but its measurement has always depended on the experimenter's knowledge of sources of organization present in the stimulus list. It is well known, however, that subjects are quite capable of memorizing materials which are not organized in any obvious manner, or for which the sources of organization cannot be readily specified. Thus, if the concept of organization is to have more general applicability, it seems necessary to develop methods that permit observation and quantification of organization as a response measure independent of the characteristics of the stimulus material. This is what the present paper attempts to do.

Behavioral manifestations of the hypothesized organizing process can best be studied under conditions where the order in which the subject recalls items is free to vary. Items which are organized into a single unit would then be expected to occur in close temporal contiguity in subject's recall. Repeated occurrences of such sequential patterns would indicate the existence of the organizing process, and the extent to which this occurs would provide an estimate of the degree of organization. Thus a quantitative analysis of sequential dependencies among items in subject's free recall on successive trials would yield measures of a response variable closely related to the hypothetical organizing process, independently of quantity of recall.

The method proposed in this paper is derived from information theory (Shannon & Weaver, 1949), and is quite similar to Miller and Frick's (1949) "index of behavioral stereotypy." It provides for a measure of sequential redundancy in repeated ordered samples of a set of items. As the experimental paradigm used here involves presentation of material completely free from any sequential redundancy, such redundancy or organization in subject's recall cannot be attributed to the input material. Rather, it is the subject who imposes a certain degree of organization on the material. For this reason such organization is called subjective organization (SO). In information theory terms, SO is "noise" produced by the channel, that is the subject; it is information in the output not found in the input (Miller, 1953).

MEASUREMENT OF SUBJECTIVE ORGANIZATION

Suppose we select L words, any words, and construct L different orders of these words such that each word appears in each of the L serial positions once and is preceded and followed by each other word just once. This can be done easily if $L + 1$ is a prime number. We then represent each of these L ordered lists to the subject on L separate trials. Words are shown one at a time on the memory drum. After each trial the subject is asked to record all the words that he remembers from the list, in any order he wishes. After all L ordered lists have thus been presented we have L recall records from the subject, one

TABLE 1

A SAMPLE RECALL MATRIX FOR AN INDIVIDUAL SUBJECT

*n*th word	(*n*+1)th word																	
	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	n_i
x_0		1	1	2	2	1	3				2	1	1			1	1	16
x_1							1		1		9	1			2			14
x_2	1	2			1	2						2			2	2		12
x_3	2	1			1	2	1	2				2	1			1		13
x_4	1	2		1			1	5				3		1	1			15
x_5	1						2		1	8				1		2		15
x_6	1	1		1					1			1		11				16
x_7	1		1		2								1		7			12
x_8	2	2				3					2		1			4		14
x_9		1			1	1			10						1	1		15
x_{10}		1	1		1	1	3			1		1	2			1	1	13
x_{11}	1				6	1	2										4	14
x_{12}	2	1						2				1		1	1	1	4	13
x_{13}		2		9		1	2			1		1						16
x_{14}	1		9					1					2	1				14
x_{15}	2				1	3	1		1	5				1				14
x_{16}	1							2				1	5			1		10
n_j	16	14	12	13	15	15	16	12	14	15	13	14	13	16	14	14	10	$N = 236$

Note.—Recall is pooled over 16 trials.

for each trial. We then tabulate the frequency of all pairs of adjacent recall responses in a matrix consisting of $L + 1$ rows and $L + 1$ columns, and compute the measure of SO from the data in the matrix.

Table 1 shows a sample recall (output) matrix for a case where $L = 16$ and where the subjects recall responses are pooled for all 16 trials. Symbols x_1, x_2, . . . x_{16} identifying the rows refer to the 1st, 2nd, . . . 16th word in the *n*th position in the subject's recall list; symbols y_1, y_2 . . . y_{16} identifying the columns stand for the same words in the $(n + 1)$th position in the subject's recall list. Thus x and y represent successive positions in the subject's recall, and subscripts identify the contents of these positions, i.e., words. Symbol x_0 stands for the blank position (no word) immediately preceding the first word in the subject's recall list, and y_0 refers to the blank position immediately following the last word in the same list. The entries in the cells of the matrix show the frequency with which word i was followed by word j in the subject's recall on all 16 trials; their numerical values are symbolized as n_{ij}. Entries in the row labelled x_0 show the frequencies with which each word j stood at the beginning of the subject's recall lists on 16 trials, and similarly entries in the column y_0 show frequencies with which each word i was in the last position of the recall lists on these trials. Zeros have been omitted in the matrix. For instance, $n_{06} = 3$, which means that Word Number 6 was in the first position (followed no other word) in the subject's recall on three trials; $n_{13} = 0$, indicating that Word Number 3 never followed Word Number 1 during the 16 trials, and so forth. The numerical values of marginal totals of rows, n_i, show how many times each word i appeared in the subject's recall on all 16 trials. Marginal totals for corresponding rows and columns are of course identical. It is also to be noted that there are no entries in the cells along the main diagonal, from upper left to lower right, since normally no word follows itself in recall.

Given a recall matrix such as the one in Table 1, a number of different measures of second-order sequential organization could be used. The measure of sequential organization adopted in the present paper is essentially a measure of redundancy in a sequence of events when the probabilities of successive pairs of events are known or can be estimated:

$$C_x(y) = 1 - \frac{H_x(y)}{\max H_x(y)} \qquad [1]$$

$H_x(y)$ here refers to the amount of information (uncertainty) in the event y when the preceding event x is known, and *max* $H_x(y)$ stands for the maximum value of such uncertainty. This is Miller and Frick's (1949) second-order index of behavioral stereotypy. The formula for computing second-order SO involves a minor modification of the above formula,[2] simplifying calculations in addition to being somewhat more appropriate to the free recall learning situation. This formula is as follows:

$$\mathrm{SO} = \frac{\sum_{i,j} n_{ij} \log n_{ij} - \min \sum_{i,j} n_{ij} \log n_{ij}}{\sum_{i} n_i \log n_i - \min \sum_{i,j} n_{ij} \log n_{ij}} \qquad [2]$$

where n_{ij} represents the numerical value of the cell in the ith row and jth column, n_i represents the marginal total of the ith row, and $\min \sum_{i,j} n_{ij} \log n_{ij}$ represents the minimum value that $\sum_{i,j} n_{ij} \log n_{ij}$ can assume.

In cases where no marginal total n_i exceeds the length of the list L, $\min \sum_{i,j} n_{ij} \log n_{ij}$ is equal to 0, and Formula 2 reduces to the following relatively simple expression:

$$\mathrm{SO} = \frac{\sum_{i,j} n_{ij} \log n_{ij}}{\sum_{i} n_i \log n_i} \qquad [3]$$

In this formula $\sum_{i} n_i \log n_i$ can be regarded as a measure of maximum organization, $\sum_{i,j} n_{ij} \log n_{ij}$ represents the actual organization, and SO is simply a measure of actual organization relative to the maximum. Thus SO can assume all values between zero and unity, the former expressing the complete absence of second-order sequential organization, as in the case of the ordered stimulus lists presented to the subject, and the latter the maximum degree of such organization, as in the case of an imaginary subject who recalls all words on all trials exactly in the same order. SO for the data in Table 1, incidentally, is .397.

So far we have considered only second-order sequential organization,

---

[2] *Max* $H_x(y)$ in Miller and Frick's index of behavioral stereotypy is defined as *max* $H(x,y) - max\ H(x)$. Thus it is based on the assumption that all responses are equally frequent. For the present problem, however, it may be somewhat more meaningful to compute *max'* $H_x(y)$ on the basis of the *given* frequency of individual responses, i.e., as determined by subject's recall:

$$max'\ H_x(y) = max\ H\ (x,y) - H\ (x)$$

If *max'* $H_x(y)$ is substituted for *max* $H_x(y)$ in Formula 1, that formula becomes mathematically equivalent with Formula 2, the measure of subjective organization as used in this paper.

based on pairs of successive responses. The method can be easily extended to higher-order dependencies, even though the computational labors involved would become quite prohibitive. However, estimates of higher-order organization can be obtained more easily. For instance, instead of setting up a three dimensional matrix for calculating third-order SO, we can tabulate data in a two dimensional matrix with nth items as rows and $(n+2)$th items as columns. In this case we also need two rows for blank spaces preceding the first word (x_0 and x_{0-1}), and two columns for blank spaces following the last word in each recall list (y_0 and y_{0+1}). This procedure amounts to collapsing the three dimensional matrix along the $(n+1)$th dimension. That is, we are computing an SO measure based on pairs of responses *separated* by one response. Such an SO is labeled SO (Lag 1), and it does provide an estimate of the third-order SO. In a similar manner, still higher order dependencies can be evaluated. SO (Lag 2) is an estimate of the fourth-order sequential organization, SO (Lag 3) is an estimate of the fifth-order SO, and so forth.

SO can be calculated for any block of trials, usually successive trials. Thus it is sometimes useful to indicate the size of the block on which the measure is based. For instance, SO (Blks = 3) means SO calculated from blocks of three successive trials. SO for a single trial is not meaningful, as subjective organization is defined in terms of the subject's tendency to recall words in the same order on successive trials; in fact SO for a single trial, when calculated according to the formulas given, turns out to be 0, provided that the subject has not repeated any pairs of words.

DEMONSTRATION EXPERIMENT: SUBJECTIVE ORGANIZATION AND PERFORMANCE

The following simple experiment will serve as an illustration of how the method works, and the data from the experiment will provide tentative answers to three questions which can be raised in the context of the discussion of the concept of organization in the first section. The questions are as follows: First, do the subjects actually organize "unrelated" words when instructions are simply to recall as many words as they can? Second, does this subjective organization , if found in the data, increase with repetition of the material? Third, is there a systematic relation between subjective organization and more traditional measures of learning, such as the number of words correctly recalled?

Procedure

A list of 16 English words was adopted from another verbal learning experiment (List III in Tulving & Thornton, 1959). All words were disyllabic nouns, consisting of five, six, or seven letters. The words, here listed alphabetically, were as follows: ACCENT, BARRACK, DRUMLIN, FINDING, GARDEN, HOYDEN, ISSUE, JUNGLE, LAGOON, MAXIM, OFFICE, POMADE, QUILLET, TREASON, VALLEY, WALKER. Sixteen different sequences of these words were constructed such that, considering any block or all 16 orders of the stimulus list, there was no second-order or higher-order redundancy in the lists.

Sixteen female undergradute students enrolled in the introductory psychology courses at the University of Toronto served as subjects. Their median age was 19 years. None of the subjects had previously participated in any verbal learning experiments.

Each subject was tested individually. In the instructions the subject was told that her task was to learn a list of 16 words, two-syllable English nouns, that would be presented to her on 16 separate trials. At the end of each trial she was to write down as many words from the list as she remembered. Then she would be shown the same words again, in a different order, and at the end of the trial she would again write down all the words that she was able to recall. The subject was also told that the order in which she recalled the words did not matter, her task was simply to recall as many of the words on each trial as she could.

The 16 different orders of the stimulus list were typed on white paper in lower case letters and presented to the subject on a memory drum. The rate of exposure was one word per second. At the end of the trial the subject was given 90 seconds to record her recall on sheets of paper lined with 16 consecutively numbered lines.

The order in which the 16 ordered lists were presented to the subjects was systematically counterbalanced among subjects. The method used for assigning the 16 orders of the stimulus list to the 16 subjects on the 16 trials was the same as that used in assigning the 16 words to 16 serial positions in 16 orders of the stimulus list.

Results

Two kinds of data are of interest in this experiment, performance (P) defined in terms of frequency of correct recall, and subjective organization (SO). Extralist intrusions and misspelled words from the list were ignored in computing both P and SO scores.[3] Repetitions of list words within a given trial were included in the recall matrix and thus entered the SO score, but not the P score. The mean number of extralist intrusions, together with misspelled words, was 0.13 per trial, the mean number of repetitions of list words was 0.09 per trial.

Second-order SO scores were calculated for successive blocks of two trials (SO, Blks = 2), for running blocks of three trials (SO, Blks = 3), and for the total block of 16 trials

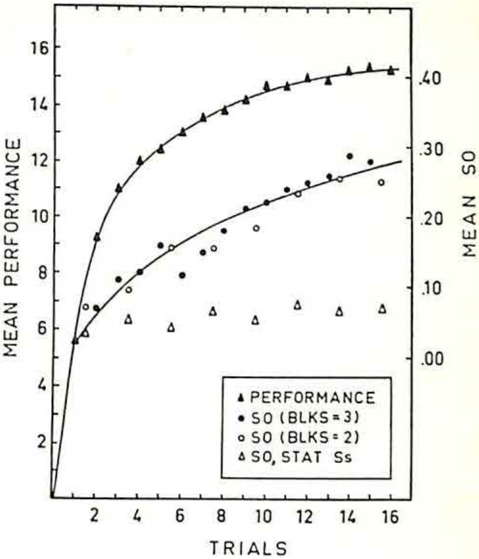


FIG. 1. Mean performance (upper curve) and mean SO (lower curve) as a function of trials. (Values of performance are to be read from the left ordinate, SO from the right ordinate. Mean SO for blocks of two trials from 16 statistical subjects are shown for comparison, open triangles. For further explanation see the text.)

(SO, Blks = 16). SO (Blks = 2) scores were also computed for 16 statistical subjects.[4]

The effect of repetition on both P and SO is summarized graphically in Figure 1. The values of P are to be read from the left, the values of SO from the right hand ordinate. The abscissa represents trials. The upper curve is the typical learning curve, showing performance as a function of trials. The lower curve, fitted by inspection to data points representing SO (Blks = 3), could be called an

---

[3] Such extralist intrusions are of interest in their own right and may provide valuable information about the organizing process. In order to handle them in the present situation, the recall matrix can be simply extended. However, such intrusions can probably be better studied in different experimental situations (e.g., Deese, 1959).

[4] IBM Model 650 Electronic Data Processing Machine was used in these calculations. Thanks are due to Albert S. Bregman for constructing the program. Data from statistical subjects, whose "performance" was matched with that of individual experimental subjects, were collected as an attempt to estimate the amount of organization occurring by chance, since the sampling distribution of SO, under the null hypothesis of no organization, is not known.

"organization" curve. The SO for each block is plotted against the middle trial of the block. Thus there are 14 SO (Blks = 3) scores. SO scores for successive blocks of size two are also shown for both the real and the statistical subjects, plotted against the midpoints of the trial blocks.

These data clearly demonstrate that subjects do in fact organize their recall sequentially even in the absence of such sequential organization in stimulus lists, and that this organization increases systematically with repeated exposures to, and recall of the material. Thus, in an experimental situation such as the present one, repetition has two parallel effects: increasing frequency of recall is accompanied by an increasingly tighter sequential organization.

The slopes of the two curves depicted in Figure 1 seem to be different. When mean P scores are plotted against mean SO (Blks = 3) for the 14 trials, the curvilinearity of the relation is obvious. However, the relation becomes quite linear when a logarithmic transformation is applied to the mean SO scores, suggesting that, as a first approximation, increase in performance is proportional to increase in log SO. The product-moment correlation between mean P and log mean SO (Blks = 3) for Trials 2 to 15 was found to be +.96.

Estimates of higher-order SO were obtained in the form of SO scores for Lags 1, 2, and 3, for the total block of 16 trials. The mean SO (Blks = 16) scores for these lags and their standard deviations are shown in Table 2, together with comparable data from the statistical subjects for Lags 0 and 3. Even though mean SO scores seem to decrease with increasing lags, it is interesting to note that SO (Lag 3, Blks = 16) scores are higher than the same scores from the statistical subjects. The median test yielded a chi-square of 15.12, which is highly significant.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF SO (Blks = 16) FOR VARIOUS LAGS

Group	Lag 0	Lag 1	Lag 2	Lag 3
Experimental subjects				
Mean	.292	.224	.223	.216
Standard deviation	.054	.045	.034	.035
Statistical subjects				
Mean	.180			.169
Standard deviation	.017			.012

Note.—N = 16 for each group.

TABLE 3

INTERCORRELATIONS AMONG SO (Blks = 16) MEASURES FOR LAGS 0, 1, 2, AND 3; AND CORRELATIONS BETWEEN THESE SO SCORES AND MEAN P SCORES ON TRIALS 1 TO 8, 9, TO 16, AND 1 TO 16 FOR INDIVIDUAL SUBJECTS

Measures	SO			
	Lag 0	Lag 1	Lag 2	Lag 3
SO (Lag 1)	+.86			
SO (Lag 2)	+.62	+.85		
SO (Lag 3)	+.58	+.74	+.84	
P (Trials 1 to 8)	+.45	+.32	+.24	+.01
P (Trials 9 to 16)	+.78	+.61	+.34	+.19
P (Trials 1 to 16)	+.63	+.47	+.30	+.08

Note.—N = 16.

The intercorrelations among SO (Blks = 16) scores for Lags 0, 1, 2, and 3, as well as correlations between these SO scores and several average P measures are shown in Table 3. These correlations are based on individual data from 16 subjects.

Certain orderly relations appear in Table 3. First, the correlations between any two adjacent SO measures on the lag dimension are reasonably high, all three listed in the table being approximately +.85. Second, with increasing distance on the lag dimension between any two SO measures the correlations decrease. Thus, for example, SO (Lag 0) has a correlation of +.85 with that for Lag 1, +.62 with Lag 2, and +.58 with Lag 3. Third, correlations between P and SO decrease with increase in the lag of the SO measure. The most important conclusion to be drawn from these data is that, within the limits of the present method, the second-order SO is about as useful a measure of sequential organization as that based on any combination of different orders, and more useful than any other single higher-order SO measure.

Positive correlations between SO and P scores in Table 3, where correlation is done on data from 16 subjects, support the observation of the same relation over trials. It is also interesting to note that on the basis of a single estimate of organization for each subject, that based on the total block of 16 trials, SO accounts for a larger amount of variance in the P variable for Trials 9 to 16 than for Trials 1 to 8.

All the above findings relate to organization as an intrasubject phenomenon. Commonality of organization between subjects, however, can also be investigated by using the same general method. That there might be such commonality became apparent in the inspection of recall data from different subjects. Recall matrices from individual subjects showed that quite often the patterning of response sequences was similar for many subjects. This phenomenon was investigated in two ways.

First, the recall data from all 16 subjects were pooled for all 16 trials and entered into an "intersubject recall matrix." If all subjects organized their recall differently, the distribution of cell entries in the group recall matrix would not be significantly different from the distribution under conditions where only chance factors are operating. To test the significance of the apparent deviation of the data in this matrix from chance distribution, the information function $T(x,y)$ was computed. It was found to be 0.352 bits which is significant at better than .01 level, using the method suggested by Miller (1955). This finding thus confirms the casual observation that there is a certain degree of commonality in the subject's recall.

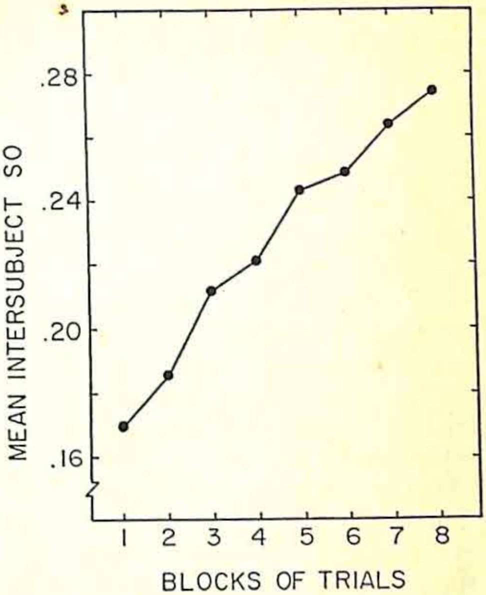


FIG. 2. Mean intersubject SO scores for 16 subjects on eight successive blocks of two trials.

The second method used to explore intersubject organization was as follows. For a single trial data were pooled in a recall matrix from all 16 subjects. For each such intersubject recall matrix SO was computed. This procedure results in 16 intersubject SO measures, one for each trial. These data are shown in Figure 2. Intersubject SO measures have been averaged for blocks of two trials, in order to smooth the curve. In view of the orderly increase in the intersubject SO measure with repetition, it would seem safe to conclude that the commonality of organized recall sequences between subjects increases with repetition of the material.

DISCUSSION

All three questions posed at the beginning of the experiment have been answered in the affirmative. It seems that the subjects do impose a sequential structure on their recall, that this subjective organization increases with repeated exposures and recall of the material, and that there is a positive correlation between organization and performance. None of these findings is very surprising, and all are in good agreement with related experimental evidence. Bousfield and his associates, as well as other investigators, have amply demonstrated that there is a strong tendency in subjects to recall randomly presented material in sequences of related words or clusters (Bousfield, 1953; Cohen & Bousfield, 1956; Jenkins & Russell, 1952). It has also been shown that repeated presentations of the stimulus word list increases this clustering tendency (Bousfield & Cohen, 1953), and that there is a direct relation between the degree of clustering and the amount of recall (Bousfield & Cohen, 1955; Bousfield, Cohen, & Whitmarsh, 1958; Jenkins, Mink, & Russell, 1958; Sakoda, 1956). The findings of the present experiment extend the domain of these phenomena of organization from one trial recall and experimentally organized materials, as used in all the above experiments, to a learning situation, with several successive trials, and experimentally unorganized materials.

The finding that the organizing effects, well demonstrated in experiments on clustering, can be experimentally assessed even in case of unrelated words, is quite encouraging. Perhaps paradoxically this suggests that a list of completely *unrelated* words is probably as fictional as is a truly nonsensical nonsense syllable. It is for this reason that reference has been made throughout this paper to "unrelated" words, referring only to the fact that words are experimentally unselected as to their meaning.

The present method of quantifying sequential organization in free recall learning constitutes only a rather coarse net for capturing the process of organization in all its diversity. For one thing, the SO measure is most sensitive to rigid sequences of responses, or chunks (Miller, 1956c). It is quite likely, however, that a great deal of organization occurs in the form of clusters (Bousfield, 1953), in which items have no fixed order. This would tend not only to attenuate measures of SO, but also, through increased variability of SO, depress correlations with the performance variable.

It would be interesting, in view of Miller's unitization hypothesis, to look at organization in terms of its units. In many cases subjects undoubtedly impose a hierarchical organization on recall, or use a Plan (Miller, Galanter, & Pribram, 1960). In other cases, particularly when the list to be learned is short and some coding device such as the alphabet can be employed, nonhierarchical organization may also occur. The present method does not lend itself to an identification of various strategies adopted by subjects, even though casual inspection suggests many diverse sources of organization: associative grouping, conceptual categories, assonance, grouping in terms of familiarity of items, and so forth.

The method provides only an overall measure of organization. This measure is related to the size of the organizing units in a systematic manner, but the relation cannot be easily specified. The two extremes of the function, however, are determined by the definition of the SO measure, and the transition from one to the other must be gradual. In one case the units of organization are individual items in the list, resulting from previous response integration (Mandler, 1954) outside the laboratory. In this case the probability of two or more items being recalled consistently together is no greater than what might be expected by chance, and the numerical value of SO is approximately the same as that of statistical subjects. On the other extreme, the whole set of items is organized into one unit, and SO is at the maximum.

Assuming a systematic relation between the size of the organizing units and the numerical value of SO, the finding that the second-order SO accounts for a larger proportion of variance in individual P scores than do estimates of higher-order measures makes good sense. Some units of organization contain only two items, and, given a flexible order of units in recall, these are not tapped by the third-order and higher-order measures of SO, while the second-order measure is sensitive to all units of two or more items. This interpretation is also consistent with the observed intercorrelations among SO measures for different lags. Although these measures are all based on identical sequences of responses, the correlations are less than perfect. This probably reflects the fact that distributions of size of organizing units vary from subject to subject.

What is the significance of the method and the preliminary findings for the question originally posed about the effects of repetition on performance? It seems that response variables such as SO, based on ordered classifications of data, might be quite useful in our attempts to shed more light on the nature of the acquisition process in verbal learning. And the present findings seem to have contributed to the attractiveness of Miller's unitization hypothesis (1956a) as a serious beginning of a useful theory of free recall verbal learning. Even though intuitively there seems little doubt that performance depends on organization, the correlational design used in the present study does not yet permit such a conclusion, and other kinds of experiments are necessary. Given a method of quantifying behavioral manifestations of the organizing process, it is quite possible to test many implications of the unitization hypothesis.

It is to be noted that the concept of repetition has been specified here rather loosely. There are at least two operationally distinguishable phases which should be studied separately, presentation of material and test for recall. A conceptual and experimental clarification of this important concept is clearly needed.

The observation of common recall sequences among different subjects and of increased stereotypy of such intersubject organization under conditions of practice very strongly suggests, or rather confirms the expectation, that sources of organization are discovered by subjects in the material, rather than invented idiosyncratically. Many task variables, therefore, are expected to influence both subjective organization and recall of the material, and a systematic exploration of these relations constitutes an important part of the research program designed to evaluate the tenability of the unitization hypothesis.

SUMMARY

A method for examining and quantifying sequential dependencies in the subjects' free recall of words on successive recall trials has been presented. Subjective organization as a depend-

ent variable was defined in terms of the subject's tendency to recall items in the same order on different trials in the absence of any experimentally manipulated sequential organization among items in the stimulus list. In a preliminary experiment it was found that the subject's recall behavior manifests such subjective organization, that this organization increases with repetition, and that there is a positive correlation between organization and performance. These data were discussed with reference to the problem of the role of repetition in free recall verbal learning.

REFERENCES

BOLLES, R. C. The effect of altering the middle of the list during serial learning. *Amer. J. Psychol.*, 1959, **72**, 577–580.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

BOUSFIELD, W. A., & COHEN, B. H. The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. *J. Psychol.*, 1953, **36**, 67–81.

BOUSFIELD, W. A., & COHEN, B. H. The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. *J. gen. Psychol.*, 1955, **52**, 83–95.

BOUSFIELD, W. A., COHEN, B. H., & WHITMARSH, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychol. Rep.*, 1958, **4**, 39–44.

COFER, C. N. A study of clustering in free recall based on synonyms. *J. gen. Psychol.*, 1959, **60**, 3–10.

COHEN, B. H., & BOUSFIELD, W. A. The effects of a dual-level stimulus-word list on the occurrence of clustering in recall. *J. gen. Psychol.*, 1956, **55**, 51–58.

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22.

JENKINS, J. J., MINK, W. D., & RUSSELL, W. A. Associative clustering as a function of verbal association strength. *Psychol. Rep.*, 1958, **4**, 127–136.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–821.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

MILLER, G. A. What is information measurement? *Amer. Psychologist*, 1953, **8**, 3–11.

MILLER, G. A. Note on the bias of information estimates. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955.

MILLER, G. A. Human memory and the storage of information. *IRE Trans. inform. Theory*, 1956, **2**, 129–137. (a)

MILLER, G. A. Information and memory. *Scient. American*, 1956, **195**, 42–46. (b)

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–96. (c)

MILLER, G. A., & FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

ROTHKOPF, E. Z., & COKE, E. U. The prediction of free recall from word association measures. *J. exp. Psychol.*, 1961, **62**, 433–438.

SAKODA, J. M. Individual differences in correlation between clustering and recall of meaningful words. *J. gen. Psychol.*, 1956, **54**, 183–190.

SHANNON, C., & WEAVER, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.

TULVING, E., & THORNTON, G. B. Interaction between proaction and retroaction in short-term retention. *Canad. J. Psychol.*, 1959, **13**, 255–265.

(Received November 28, 1960)

Psychological Review
1962, Vol. 69, No. 4, 355-365

THE ACQUISITION OF KNOWLEDGE[1]

ROBERT M. GAGNÉ

Princeton University

The growing interest in autoinstructional devices and their component learning programs has had the effect of focusing attention on what may be called "productive learning." By this phrase is meant the kind of change in human behavior which permits the individual to perform successfully on an entire *class* of specific tasks, rather than simply on one member of the class. Self-instructional programs are designed to ensure the acquisition of capabilities of performing classes of tasks implied by names like "binary numbers," "musical notation," "solving linear equations" rather than tasks requiring the reproduction of particular responses.

When viewed in this manner, learning programming is not seen simply as a technological development incorporating previously established learning principles, but rather as one particular form of the ordering of stimulus and response events designed to bring about productive learning. It should be possible to study such learning, and the conditions which affect it, by the use of any of a variety of teaching machines, although there are few studies of this sort in the current literature (cf. Lumsdaine & Glaser, 1960). In the laboratory, the usual form taken by studies of productive learning has been primarily that of the effects of instructions and pretraining on problem solving (e.g., Hilgard, Irvine, & Whipple, 1953; Katona, 1940; Maltzman, Eisman, Brooks, & Smith, 1956).

When an individual is subjected to the situation represented by a learning program, his performance may change, and the experimenter then infers that he has acquired a new capability. It would not be adequate to say merely that he has acquired new "responses," since one cannot identify the specific responses involved. (Adding fractions, for example, could be represented by any of an infinite number of distinguishable stimulus situations, and an equal number of responses.) Since we need to have a term by means of which to refer to what is acquired as a result of responding correctly to a learning program, we may as well use the term "knowledge." By definition, "knowledge" is that inferred capability which makes possible the successful performance of a *class of tasks* that could not be performed before the learning was undertaken.

Some Initial Observations

In a previous study of programmed learning (Gagné & Brown, 1961) several kinds of learning programs were used in the attempt to establish the performance, in high school boys, of deriving formulas for the sum of n terms in a number series. Additional observations with this material led us to the following formulation: In productive learning, we are dealing with two major categories of variables. The first of these is knowledge, that is, the capabilities the individual possesses at any given stage in the learning; while the second is instructions, the content of the communications presented within the frames of a learning program.

---

[1] This study was made possible in part by funds granted by the Carnegie Corporation of New York. The opinions expressed are those of the author, and do not necessarily reflect the views of that Corporation.

In considering further the knowledge category, it has been found possible to identify this class of variable more comprehensively in the following way: Beginning with the final task, the question is asked, What kind of capability would an individual have to possess if he were able to perform this task successfully, were we to give him only instructions? The answer to this question, it turns out, identifies a new class of task which appears to have several important characteristics. Although it is conceived as an internal "disposition," it is directly measurable as a performance. Yet it is *not the same* performance as the final task from which it was derived. It is in some sense *simpler,* and it is also *more general.* In other words, it appears that what we have defined by this procedure is an entity of "subordinate knowledge" which is essential to the performance of the more specific final task.

Having done this, it was natural to think next of repeating the procedure with this newly defined entity (task). What would the individual have to know in order to be capable of doing *this* task without undertaking any learning, but given only some instructions? This time it seemed evident that there were two entities of subordinate knowledge which combined in support of the task. Continuing to follow this procedure, we found that what we were defining was a *hierarchy* of subordinate knowledges, growing increasingly "simple," and at the same time increasingly general as the defining process continued.

By means of this systematic analysis, it was possible to identify nine separate entities of subordinate knowledge, arranged in hierarchical fashion (see Figure 1). Generally stated, our hypothesis was that (*a*) no individual could perform the final task without having these subordinate capabilities (i.e., without being able to perform these simpler and more general tasks); and (*b*) that any superordinate task in the hierarchy could be performed by an individual provided suitable instructions were given, and provided the relevant subordinate knowledges could be recalled by him.

It may be noted that there are some possible resemblances between the entities of such a knowledge hierarchy and the hypothetical constructs described by three other writers. First are the habit-family hierarchies of Maltzman (1955), which are conceived to mediate problem solving, and are aroused by instructions (Maltzman et al., 1956). The second are the "organizations" proposed by Katona (1940), which are considered to be combined by the learner into new knowledge after receiving certain kinds of instructions, without repetitive practice. The third is Harlow's (1949) concept of learning set. Harlow's monkeys acquired a general capability of successfully performing a class of tasks, such as oddity problems, and accordingly are said to have acquired a learning set. There is also the suggestion in one of Harlow's (Harlow & Harlow, 1949) reports that there may be a hierarchical arrangement of tasks more complex than oddity problems which monkeys can successfully perform. Since we think it important to imply a continuity between the relatively complex performances described here and the simpler ones performed by monkeys, we are inclined to refer to these subordinate capabilities as "learning sets."

REQUIREMENTS OF THEORY

If there is to be a theory of productive learning, it evidently must deal with the independent variables that can be identified in the two major categories of instructions and subordinate capabilities, as well as with their interactions, in bringing about changes in human performance.

Instructions

Within a learning program, instructions generally take the form of sentences which communicate something to the learner. It seems possible to think of such "communication" as being carried out with animals lower than man, by means of quite a different set of experimental operations. Because of these communications, the human learner progresses from a point in the learning sequence at which he can perform one set of tasks to a point at which he achieves, for the first time, a higher level learning set (class of tasks). What functions must a theory of knowledge acquisition account for, if it is to encompass the effects of instructions? The following paragraphs will attempt to describe these functions, not necessarily in order of importance.

First, instructions make it possible for the learner to *identify the required terminal performance* (for any given learning set). In educational terms, it might be said that they "define the goal." For example, if the task is adding fractions, it may be necessary for the learner to identify 15¾ as an adequate answer, and $^{63}/_4$ as an inadequate one.

Second, instructions bring about proper *identifications of the elements* of the stimulus situation. For example, suppose that problems are to be presented using the word "fraction." The learner must be able to identify ⅖ as a fraction and .4 as not a fraction. Or, he may have to identify Σ as "sum of," and n as "number." Usually, instructions establish such identifications in a very few repetitions, and sometimes in a single trial. If there are many of them, differentiation may require several repetitions involving contrasting feedback for right and wrong responses.

A third function of instructions is to establish *high recallability* of learning sets. The most obviously manipulable way to do this is by repetition. However it should be noted that repetition has a particular meaning in this context. It is not exact repetition of a stimulus situation (as in reproductive learning), but rather the presentation of additional examples of a class of tasks. Typically, within a learning program, a task representing a particular learning set is achieved once, for the first time. This may then be followed by instructions which present one or more additional examples of this same class of task. "Variety" in such repetition (meaning variety in the stimulus context) may be an important subvariable in affecting recallability. Instructions having the function of establishing high recallability for learning sets may demand "recall," as in the instances cited, or they may on other occasions attempt to achieve this effect by "recognition" (i.e., not requiring the learner to produce an answer).

The fourth function of instructions is perhaps the most interesting from the standpoint of the questions it raises for research. This is the "guidance of thinking," concerning whose operation there is only a small amount of evidence (cf. Duncan, 1959). Once the subordinate learning sets have been recalled, instructions are used to promote their application to (or perhaps "integration into") the performance of a task that is entirely new so far as the learner is concerned. At a minimum, this function of instructions may be provided by a statement like "Now put these ideas together to solve this problem;" possibly this amounts to an attempt to establish a *set*. Beyond this, thinking may be guided by suggestions which progressively limit the range of hypotheses entertained by the learner, in such a way as to decrease the number of incorrect solutions he considers (cf. Gagné & Brown, 1961; Katona, 1940). Within a typical learning program, guidance of thinking is employed after identification of terminal performance and of stimulus elements have been completed, and after high recallability of relevant learning sets has been ensured. In common sense terms, the purpose of these instructions is to suggest to the learner "how to approach the solution of a new task" without, however, "telling him the answer."

Obviously, much more is needed to be known about the effects of this variable, if indeed it is a single variable. Initially, it might be noted that guidance of thinking can vary in *amount;* that is, one can design a set of instructions which say no more than "now do this new task" (a minimal amount); or, at the other end of the scale, a set of instructions which in effect suggest a step-by-step procedure for using previously acquired learning sets in a new situation.

Subordinate Capabilities: Learning Sets

When one begins with the performance of a particular class of tasks as a criterion of terminal behavior, it is possible to identify the subordinate learning sets required by means of the procedure previously described. The question may be stated more exactly as, "What would the individual have to be able to do in order that he can attain successful performance on this task, provided he is given only instructions?" This question is then applied successively to the subordinate classes of tasks identified by the answer. "What he would have to be able to do" is in each case one or more performances which constitute the denotative definitions of learning sets for particular classes of tasks, and totally for the entire knowledge hierarchy.

A theory of knowledge acquisition must propose some manner of functioning for the learning sets in a hierarchy. A good possibility seems to be that they are mediators of positive transfer from lower-level learning sets to higher-level tasks. The hypothesis is proposed that specific transfer from one learning set to another standing above it in the hierarchy will be zero if the lower one cannot be recalled, and will range up to 100% if it can be.

In narrative form, the action of the two classes of variables in the acquisition of knowledge is conceived in the following way. A human learner begins the acquisition of the capability of performing a particular class of tasks with an individual array of relevant learning sets, previously acquired. He then acquires new learning sets at progressively higher levels of the knowledge hierarchy until the final class of tasks is achieved. Attaining each new learning set depends upon a process of positive transfer, which is dependent upon (*a*) the recall of relevant subordinate learning sets, and upon (*b*) the effects of instructions.

EXPERIMENTAL PREDICTIONS AND RESULTS [2]

Using the procedure described, we derived the knowledge hierarchy depicted in Figure 1 for the final task of "deriving formulas for the sum of *n* terms in number series."

As mentioned previously, it contained nine hypothesized learning sets. (The final row of circled entities will

---

[2] The author is grateful to Bert Zippel, Jr., for assistance in the preparation of learning program materials and in the collection of a portion of the data.

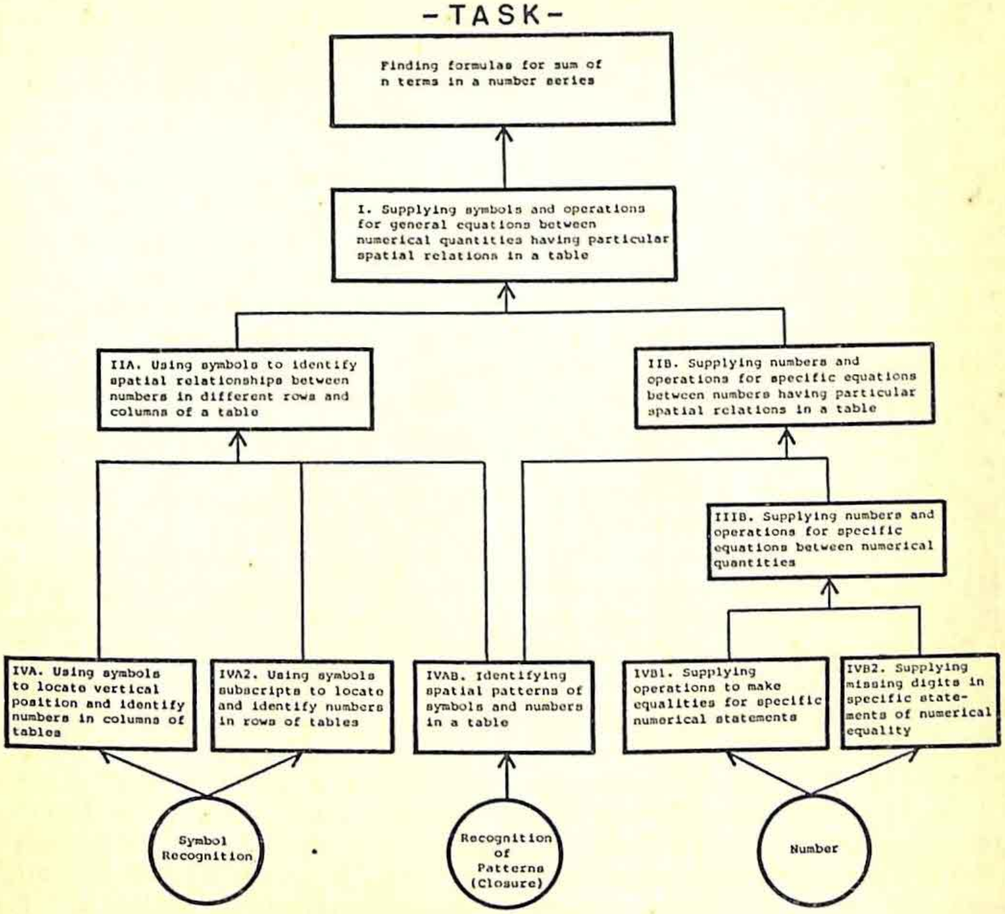

-TASK-
Finding formulas for sum of n terms in a number series
I. Supplying symbols and operations for general equations between numerical quantities having particular spatial relations in a table
IIA. Using symbols to identify spatial relationships between numbers in different rows and columns of a table
IIB. Supplying numbers and operations for specific equations between numbers having particular spatial relations in a table
IIIB. Supplying numbers and operations for specific equations between numerical quantities
IVA. Using symbols to locate vertical position and identify numbers in columns of tables
IVA2. Using symbols subscripts to locate and identify numbers in rows of tables
IVAB. Identifying spatial patterns of symbols and numbers in a table
IVB1. Supplying operations to make equalities for specific numerical statements
IVB2. Supplying missing digits in specific statements of numerical equality
Symbol Recognition
Recognition of Patterns (Closure)
Number

FIG. 1. Hierarchy of knowledge for the task of finding formulas for the for the sum of n terms in a number series.

be discussed later.) Each of these subordinate knowledges can be represented as a class of task to be performed.

Measuring Initial Patterns of Learning Sets

It is predicted that the presence of different patterns of learning sets can be determined for individuals who are unable to perform a final task such as the one under consideration. To test this, we administered a series of test items to a number of ninth-grade boys. These items were presented on 4″ × 6″ cards, and the answers were written on specially prepared answer sheets. This particular method was used in order to make testing continuous with the administration of a learning program to be described hereafter. Each test item was carefully prepared to include instructions having the function of identification of terminal performance and of elements of the stimulus situation.

Beginning with the final task, the items were arranged to be presented in the order I, IIA, IVA, IVA2, IVAB, IIB, IIIB, IVB1, and IVB2. For any given subject, the sequence of testing temporarily stopped at the level at which successful performance was first reached, and a learning program designed to foster achievement at the next higher level (previously failed) was administered. This program and its results will shortly be described. Following this, testing on the remaining learning set tasks was undertaken in the order given. The possibility of effects of the learning program on the performance of these lower-level learning sets (not specifically practiced in the learning) is of course recognized, but not further considered in the present discussion.

A particular time limit was set for each test item, at the expiration of which the item was scored as failed. If a wrong answer was given before this time limit, the subject was told it was wrong, and encouraged to try again; if the correct answer was supplied within the time limit, the item was scored as passed. It is emphasized that these time limits, which were based on preliminary observations on other subjects with these tasks, were *not* designed to put "time pressure" on the subjects, nor did they appear to do so.

The patterns of success achieved on the final task and all subordinate learning set tasks, by all seven subjects, are shown in Table 1. The subjects have been arranged in accordance with their degree of success with all tasks, beginning with one who failed the final task but succeeded at all the rest. Several things are apparent from these data. First of all, it is quite evident that there are quite different "patterns of capability" with which individuals approach the task set by the study. Some are unable to do a task like IIA (see Figure 1), others to do a task like IIB which is of course quite different. Still others are unable to do either of these, and in fact cannot perform successfully a task like IIIB. All seven of these subjects were able to perform IV-level tasks successfully, although in preliminary observations on similar tasks we found some ninth-grade boys who could not.

Second, the patterns of pass and fail on these tasks have the relationships predicted by the previous discussion. There are no instances, for example, of an individual who is able to perform what has been identified as a "higher-level" learning set, and who then shows himself to be unable to perform a "lower-level" learning set related to it.

If learning sets are indeed essential for positive transfer, the following consequences should ensue:

1. If a higher-level learning set is passed (+), *all* related lower-level tasks must have been passed (+).

TABLE 1

PATTERN OF SUCCESS ON LEARNING SET TASKS RELATED TO THE FINAL NUMBER SERIES TASK FOR SEVEN NINTH GRADE BOYS

Subject	Task									
	Final	I	IIA	IIB	IIIB	IVA1	IVA2	IVAB	IVB1	IVB2
WW	−	+	+	+	+	+	+	+	+	+
WC	−	+	+	+	+	+	+	+	+	+
PM	−	−	−	+	+	+	+	+	+	+
GR	−	−	−	+	+	+	+	+	+	+
DJ	−	−	−	−	+	+	+	+	+	+
JR	−	−	−	−	−	+	+	+	+	+
RH	−	−	−	−	−	+	+	+	+	+

Note.—+ = Pass; − = Fail.

TABLE 2

PASS-FAIL RELATIONSHIP BETWEEN RELATED ADJACENT HIGHER- AND LOWER-LEVEL LEARNING SETS FOR A GROUP OF SEVEN NINTH-GRADE BOYS

Relationship examined	Number of cases with relationship				Test of relationships	
	Higher + Lower +	Higher − Lower −	Higher + Lower −	Higher − Lower +	N (1 +2 +3)	Proportion (1 +2)/(1 +2 +3)
Final Task: I	0	6	0	1	6	1.00
I: IIA, IIB	1	5	0	1	6	1.00
IIA: IVA1, IVA2, IVAB	2	0	0	5	2	1.00
IIB: IIIB	3	2	0	2	5	1.00
IIIB: IVAB, IVB1, IVB2	5	0	0	2	5	1.00

Note.—+ = Pass; − = Fail.

2. If *one or more* lower-level tasks have been failed (−), the related higher-level tasks must be failed (−).

3. If a higher-level task is passed (+), *no* related lower-level tasks must have been failed (−).

4. If a higher-level task has been failed (−), related lower-level tasks may have been passed (+). The absence of positive transfer in this case would be attributable to a deficiency in instructions, and does not contradict the notion that lower-level sets are essential to the achievement of higher-level ones.

The relationships found to exist in these seven subjects are summarized in Table 2, where each possible higher–lower-level task relationship possible of testing is listed in the left-hand column. It will be noted that there are several relationships of the type higher (−), lower (+), as listed in Column 5. These provided no test of the hypothesis regarding hierarchical relations among learning sets. The instances in the remaining columns do, however. The ++ and −− instances are verifying, whereas +− instances would be nonverifying. As the final column indicates, the percentage of verifying instances is in all cases 100%.

Effects of Learning Program Administration

If the characteristics of instructions as previously described are correct, it should be possible to construct a learning program which can be begun for each individual at the point of his lowest successful learning set achievement, and bring him to successful achievement of the final class of tasks. Briefly, its method should be to include frames which have the functions of (*a*) insuring high recallability of relevant learning sets on which achievement has been demonstrated; (*b*) making possible identifications of expected performance and of new stimuli, for each newly presented task; and (*c*) guiding thinking so as to suggest proper directions for hypotheses associating subordinate learning sets with each new one.

A program of this sort was administered to each of the seven ninth-grade boys, beginning at the level at which he first attained success on learning set tasks (Table 1). This was done by means of a simple teaching machine consisting of a visible card file clipped to a board mounted at a 40° angle to the learner's table, and containing material typed on 4″ × 6″ cards. He wrote his answer to successive frames on a numbered answer sheet, then flipped over the card to see the correct answer on the back. He was instructed that if his answer was wrong, he should flip the card back, and read the frame again until he could "see" what the right answer was.

After completing the instructional portion of the program for each learning set, the learner was again presented with the identical test-item problem he had tried previously and failed. If he was now able to do it correctly, he was given five additional items of the same sort to perform, and then taken on by instructions to another learning set in either a coordinate or higher-level position in the hierarchy. This process was continued through the performance of the final task.

The data collected in this way yield pass-fail scores on each test item (representing a particular member of a class of tasks) *before* the administration of the learning program, and similar scores on the same item *after* learning. It is recognized that for certain experimental purposes, one would wish to have a different, matched, task for the test given after learning, to control for the effects of "acquaintance" during the first test. Since this study had an exploratory character, such a control was not used this time. However, it should be clearly understood that the first experience with these test items in question, for these subjects, involved only activity terminating in failure to achieve solution. No information about the correct solutions was given.

A striking number of instances of success in achieving correct solutions to learning set tasks was found following learning as compared with before. These results are summarized in Table 3. Although for learning set IIB the percentage of success was only 50% (with two cases), there were two learning sets for which 100% success was achieved, and the percentage for all instances combined was 86%. These results provide additional evidence compatible with the idea of the knowledge hierarchy.

TABLE 3

NUMBER OF INSTANCES OF PASSING AND FAILING FINAL TASK AND SUBORDINATE LEARNING SET TASKS BEFORE AND AFTER ADMINISTRATION OF AN ADAPTIVE LEARNING PROGRAM, IN A GROUP OF SEVEN NINTH-GRADE BOYS

Task	Number failing before learning	Number of these passing after learning	Percentage success
Final Task	7	6	86
I	5	4	80
II	5	5	100
IIB	2	1	50
IIIB	2	2	100
Total	21	18	86

The learner in such a program does not "practice the final task"; he acquires specifically identified capabilities in a specified order. In as many as six out of seven cases, we were able by this means to bring learners from various levels of competence all the way to final task achievement. (It is perhaps important that the exception was JR, one of two who had most to learn.) Of course, it must be recognized that two separable causes contribute to the effects of the learning program in this study: (*a*) the correctness of the learning set analysis; and (*b*) the specific effectiveness of the instructions contained in the learning program.

IMPLICATIONS FOR INDIVIDUAL DIFFERENCES MEASUREMENT

It is evident that learning sets, as conceived in this paper, operate as "individual differences" variables, which, when suitably manipulated, also become "experimental" variables. There are some additional implications which need to be pointed out regarding the functioning of learning sets in the determination of measured individual differences.

As the process of identification of subordinate learning sets is progressively continued, one arrives at some learning sets which are very simple and general, and likely to be widespread within the population of learners for which the task is designed. Consider, for example, learning set IVB (Figure 1), which is represented by a task such as $4 \times 2 = 5+?$ If one makes a further analysis to identify a subordinate learning set for this task, the answer appears to be, "adding, subtracting, multplying, and dividing one- and two-place numbers." It is interesting to note that this is exactly the task provided by a set of factor reference tests (French, 1954) called Number. In a similar manner, the other two circled entities in the last row of Figure 1 were identified. One is Symbol Recognition (called Associative Memory by the factor researchers), and Recognition of Patterns (called Flexibility of Closure). The implication is, therefore, that these simplest tasks, identified by factor analysis techniques as common to a great variety of human performances, also function as learning sets.

The hypothesis has been proposed that learning sets mediate positive transfer to higher-level tasks. Very often, if not usually, the measurement of transfer of training implies that a second task is learned *more* rapidly when preceded by the learning of an initial task, than when not so preceded. Accordingly, it seems necessary to distinguish between expected correlations of these basic factors (at the bottom of the hierarchy) with rate of attainment of higher-level learning sets on the one hand, and correlations of these same factors with achievement of higher-level learning sets on the other.

The implications of this line of reasoning would seem to be somewhat as follows: Factors which are found by the kind of psychological analysis previously described to lie at the bottom of the knowledge hierarchy should exhibit certain predictable patterns of correlation with higher-level learning sets. They should correlate most highly with rate of attainment of the learning sets in the next higher level to which they are related, and progressively less as one progresses upwards in the hierarchy. The reason for this is simply that the rate of attainment of learning sets in a hierarchy comes to depend to an increasing extent on the learning sets which have just previously been acquired and accordingly to a decreasing extent upon a basic factor or ability. Some analogy may be drawn here with the findings of Fleishman and Hempel (1954) on motor tasks.

The expected relationships between factor test scores and achievement scores (passing or failing learning sets) throughout such hierarchies seem to require a somewhat more complex derivation. First of all, such relationships will depend upon the effectiveness of a learning program, or perhaps on the effectiveness of previous learning. If the learning program is perfectly effective, for example, and if differences in rate of attainment are ignored, everyone will pass all the learning set tasks, and the variance will accordingly be reduced to zero. Under these circumstances, then, one may expect all correlations with basic factors to be zero. However, one must consider the case in which the learning program is not perfectly effective. In such a case, the probability that an individual will acquire a new learning set, as opposed to not acquiring it, will presumably be increased to the extent that he scores high on tests of related basic abilities. If one continues to collect scores on learning set tasks of both successful achievers and those who fail, the result will presumably be an increasing degree of correlation between basic ability scores and learning set tasks as one progresses upwards in the hierarchy. The reason for this is that the size of the correlation comes to depend more and more upon variance contributed by those individuals who are successful, and less and less on that contributed by those who effectively "drop out."

The difference in expectation between the increasing pattern of correlation with achievement scores, and the decreasing pattern with measures of rate of attainment, is considered to be of rather general importance for the area of individual differences measurement. Confirmatory results have been obtained in a recent study (Gagné & Paradise, 1961) concerned with the class of tasks "solving linear algebraic equations."

DISCUSSION

The general view of productive learning implied in this paper is that it is a matter of transfer of training from component learning sets to a new activity which incorporates these previously acquired capabilities. This new activity so produced is qualitatively different from the tasks which correspond to the "old" learning sets; that is, it must be described by a different set of operations, rather than simply being "more difficult." The characteristics of tasks which make achievement of one class of task the required precursor of achievement in another, and not vice versa, are yet to be discovered. Sufficient examples exist of this phenomenon to convince one of its reality (Gagné, Mayor, Garstens, & Paradise, 1962; Gagné & Paradise, 1961). What remains to be done, presumably, is to begin with extremely simple levels of task, such as discriminations, and investigate transfer of training to tasks of greater and greater degrees of complexity, or perhaps abstractness, thus determining the dimensions which make transfer possible.

The path to research on the characteristics of instructions appears more straightforward, at least at first glance. The establishment of identifications is a matter which has been investigated extensively with the use of paired associates. The employment of instructions for this purpose may need to take into consideration the necessity for learning differentiations among the stimulus items to be identified, as well as other variables suggested by verbal learning studies. The function of inducing high recallability would seem to be a matter related to repetition of learning set tasks, and may in addition be related to time variables such as those involved in distribution of practice. As for guidance of thinking, the distinguishing of this function from others performed by instructions should at least make possible the design of more highly analytical studies than have been possible in the past.

In the meantime, the approach employed in the experiment reported here, of proceeding backwards by analysis of an already existing task, has much to recommend it as a way of understanding the learning of school subjects like mathematics and science, and perhaps others also. Naturally, every human task yields a different hierarchy of learning sets when this method of analysis is applied. Often, the relationship of higher to lower learning sets is more complex than that exhibited in Figure 1. It should be possible, beginning with any existing class of tasks, to investigate the effects of various instruction variables within the framework of suitably designed learning programs.

The major methodological implication of this paper is to the effect that investigations of productive learning must deal intensively with the kinds of variables usually classified as "individual differences." One cannot depend upon a measurement of *general* proficiency or aptitude to reveal much of the important variability in the capabilities people bring with them to a given task. Consider, for example, the seven ninth-grade boys in our study. Each of them had "had" algebra, and each of them had "had" arithmetic. There was no particularly striking relationship between their ultimate performance and their previous grades in algebra (although there is no doubt some correlation), nor between this performance and "general intelligence." But the measurement of their learning sets, as illustrated in Table 1, revealed a great deal about how they would behave when confronted with the learning program and the final task. For some, instructions had to begin, in effect, "lower down" than for others. Some could do Task 1 right away, while others could not, but could do it equally well provided they learned other things first. The methodological point is simply this: if one wants to investigate the effects of an experimental treatment on the behavior of individuals or groups who start from the same point, he would be well advised to measure and map out for each individual the learning sets relevant to the experimental task. In this way he can have some assurance of the extent to which his subjects are equivalent.

REFERENCES

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

FLEISHMAN, E. A., & HEMPEL, W. E. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19**, 239–252.

FRENCH, J. W. (Ed.) *Kit of selected tests for reference aptitude and achievement factors.* Princeton, N. J.: Educational Testing Service, 1954.

GAGNÉ, R. M., & BROWN, L. T. Some factors in the programing of conceptual learning. *J. exp. Psychol.*, 1961, **62**, 313–321.

GAGNÉ, R. M., MAYOR, J. R., GARSTENS, H. L., & PARADISE, N. E. Factors in acquiring knowledge of a mathematical task. *Psychol. Monogr.*, 1962, **76**(7, Whole No. 526.

GAGNÉ, R. M., & PARADISE, N. E. Abilities and learning sets in knowledge acquisition. *Psychol. Monogr.*, 1961, **75**(14, Whole No. 518).

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HARLOW, H. F., & HARLOW, M. K. Learning to think. *Scient. American*, 1949, **181**, 36–39.

HILGARD, E. R., IRVINE, R. P., & WHIPPLE, J. E. Rote memorization, understanding, and transfer: An extension of Katona's card-trick experiments. *J. exp. Psychol.*, 1953, **46**, 288–292.

KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

LUMSDAINE, A. A., & GLASER, R. *Teaching machines and programmed learning.* Washington, D. C.: National Education Association, 1960.

MALTZMAN, I. Thinking: From a behavioristic point of view. *Psychol. Rev.*, 1955, **62**, 275–286.

MALTZMAN, I., EISMAN, E., BROOKS, L. O., & SMITH, W. M. Task instructions for anagrams following different task instructions and training. *J. exp. Psychol.*, 1956, **51**, 418–420.

(Early publication received August 8, 1961)

Psychological Review
1962, Vol. 69, No. 4, 366–378

ON THE RELATION BETWEEN LOGIC AND THINKING

MARY HENLE[1]

Graduate Faculty of Political and Social Science,
New School for Social Research, New York

The question of whether logic is descriptive of the thinking process, or whether its relation to thinking is normative only, seems to be easily answered. Our reasoning does not, for example, ordinarily follow the syllogistic form; and we do fall into contradictions. On the other hand, logic unquestionably provides criteria by which the validity of reasoning may be evaluated. Logical forms thus do not describe actual thinking, but are concerned with the ideal, with "how we ought to think." And yet a problem seems to be concealed beneath this easy solution.

It is interesting to note that a number of the older writers on logic regarded their discipline as the science of thought. It will not be possible here to survey the vast literature, some of which bears only by implication on our problem, but some examples will illustrate this position. Cohen and Nagel (1934) summarize the older view: "An old tradition defines logic as the science of the laws of thought" (p. 18). Kant (1885) holds that "logic is a science of the necessary laws of thought, without which no employment of the understanding and the reason takes place" (p. 3). Psychology, on the other hand, supplies only the contingent, not the necessary, rules of thought. John Stuart Mill (1874) likewise views logic as comprising "the science of reasoning, as well as an art, founded on that science" (p. 18). He continues:

> Whatever has at any time been concluded justly, whatever knowledge has been acquired otherwise than by immediate intuition, depended on the observance of the laws which it is the province of logic to investigate. If the conclusions are just, and the knowledge real, those laws, whether known or not, have been observed (p. 22).

Boole (1854), as a final instance, regards "the laws of the symbols of Logic" as "deducible from a consideration of the operations of the mind in reasoning" (pp. 45–46). Boole deduces the law of contradiction, for example, from a fundamental law of thought (pp. 49–51).

More recent writers, on the other hand, have tended to reject the view that the laws of logic are those of the human understanding. Again, only a few illustrations will be given. Cohen (1944) remarks:

> That the laws of logic are not the universal laws according to which we do actually think is conclusively shown, not only by the most elementary observation or introspection, but by the very existence of fallacies (pp. 2–3).

A similar point is made by Nagel (1956):

> Little need be said in refutation of the view that logical principles formulate the "inherent necessities of thought" and are generalized descriptions of the operations of minds. Surely the actual occurrence of beliefs in logically incompatible propositions makes nonsense of the claim that the principle of noncontradiction expresses a universal fact of psychology (p. 66).

Regarding the work of Boole referred to above (*An Investigation of the Laws of Thought*), Cohen and Nagel (1934) comment: "The title is a misnomer"

---

[1] This paper was written when the author was a Fellow of the John Simon Guggenheim Memorial Foundation.

(p. 112). A more extreme statement of this point of view is by Bertrand Russell (1904):

> Throughout logic and mathematics, the existence of the human or any other mind is totally irrelevant; mental processes are studied by means of logic, but the subject-matter of logic does not presuppose mental processes, and would be equally true if there were no mental processes (p. 812).

Schiller (1930), too, holds that syllogistic reasoning "has nothing whatever to do with actual reasoning, and can make nothing of it" (p. 282). He describes the "Laws of Thought" (laws of identity, contradiction, and excluded middle) as "verbal conventions" (p. 251).

The changed point of view with regard to the relation between logical principles and the laws of thought seems to be a function of an altered intellectual climate rather than of any fundamental discoveries about the nature of reasoning. It thus seems worthwhile to reopen the question, the more so since it has implications for a number of central issues in the psychology of thinking. First, however, it may be of interest to examine a few more forms in which the present question has been raised.

Discussion of the figures of the syllogism has at times centered on their relevance to actual thinking. J. N. Keynes (1887) cites several writers who reject the fourth figure because they hold that we do not actually reason in it (pp. 230–231). Kant (1885), indeed, finds all but the first figure both useless and false; the fourth in particular he calls unnatural (pp. 84–90). Keynes, on the other hand, argues for the admission of Figure 4 on the same grounds, namely its relevance to actual thinking: "It is not actually in frequent use, but reasonings may sometimes not unnaturally fall into it" (p. 232).

Psychologists investigating reasoning processes have tended to underemphasize the role of logic in the thinking of their subjects. To illustrate, Bruner, Goodnow, and Austin (1956) suggest that

> much of human reasoning is supported by a kind of thematic process rather than by an abstract logic. The principal feature of this thematic process is its pragmatic rather than its logical structure (p. 104).

Individuals tend to prefer "empirically reasonable propositions" to logical ones (p. 104). Morgan and Morton (1944) conclude that:

> A person is likely to accept a conclusion which expresses his convictions with little regard for the correctness or incorrectness of the inferences involved. Our evidence will indicate that the only circumstance under which we can be relatively sure that the inferences of a person will be logical is when they lead to a conclusion which he has already accepted (p. 39).

Lefford (1946) states that the principles of logical inference "are techniques which are not the common property of the unsophisticated subject" (p. 144). He goes so far as to distinguish from the logical inference

> psychological inferences which may be made by the ordinary person. . . . A psychological inference is not valid or invalid except when judged as a logical inference: psychological inference is purely a fact (p. 145).

Common to all these statements by the psychologists is the assumption that logical principles are irrelevant, if not antithetical, to much actual reasoning.[2] This conclusion is derived from the high incidence of wrong inferences of subjects under test conditions, especially in the case of emotionally relevant material. A different conclusion has been drawn by Von Domarus (1944) from observation of errors in the reasoning of schizophrenic patients. This author argues, not that the reasoning of his subjects is unrelated to logic, but rather that it conforms to a logic whose laws are different from those of Aristotelian logic. This is "paralogic," which excludes the law of contradiction and "accepts identity based upon identical predicates" (p. 111). This idea has been elaborated by Arieti (1955) who sees the operation of this logic (which he calls paleologic) not only in schizophrenic and primitive thinking, as Von Domarus does, but also in dreams, in some infantile thinking, and in the transference situation in psychoanalysis.

---

[2] A different, but related, position is that represented by Dollard and Miller (1950), who hold that being logical is a learned drive. The child, it is argued, is punished for logical contradictions and absurdities, for illogical and contradictory plans. The result for most people is "a learned drive to make their explanations and plans seem logical" (p. 120). The implication is, of course, that without this specific training the individual's thinking would not be (or seem) logical.

Again it is of interest to find that the earlier writers mentioned above were equally aware of the problem of error, but viewed it in a way that was entirely compatible with their conception of logic as the science of the laws of the mind. "It is easy to see how truth is possible," writes Kant (1885, p. 44) "since in it the understanding acts according to its own essential laws." Error, however, is difficult to understand since it constitutes "a form of thought inconsistent with the understanding." Its source is thus not to be sought in the understanding itself, but rather in the "unobserved influence of the sensibility on the understanding," the sensibility being that faculty which "supplies the material for thought." Boole (1854) likewise considers that "the phaenomena of incorrect reasoning or error . . . are due to the interference of other laws with those laws of which *right* reasoning is the product" (p. 409). He reminds us that "the laws of correct inference may be violated, but they do not the less truly *exist* on this account" (p. 408). Mill (1874) is still more explicit. Discussing fallacies of ratiocination, he points out that since

> the premises are seldom formally set out,... it is almost always to a certain degree optional in what manner the suppressed link shall be filled up. . . . [A person] has it almost always in his power to make his syllogism good by introducing a false premise; and hence it is scarcely ever possible decidedly to affirm that any argument involves a bad syllogism (p. 560).

In the case of arguments consisting not of a single syllogism but of a chain of syllogisms, he considers the commonest fallacy of ratiocination to lie in a changing of the premises as the argument proceeds.

Two clearly contrasting alternatives thus present themselves: Is logic (or Aristotelian logic) largely irrelevant to the thinking process, or is it concerned with the laws of thinking? Since we will here be concerned only with deductive reasoning, we may reformulate the question more specifically in these terms. But since, as has so often been pointed out, the premises from which we reason are commonly not spelled out, since our inferences so frequently appear as enthymemes,[3] this fact must be taken into consideration. We may ask: If we know the premises—tacit as well as explicit—from which a person reasons, can we put the process in syllogistic form? Do the rules of the syllogism describe processes that the mind follows in deductive reasoning, even when the syllogistic form is not explicitly employed?

It has been shown above that the existence of error has been used as evidence for the irrelevance of logic to the actual thinking process. On the other hand, a different interpretation of error has been suggested. Since the problem of error seems to be a particularly fruitful one in which to join the issues before us, it will be taken as the context for the present discussion. Once more we may reformulate our question as follows: Do errors in deductive reasoning mean that the logical process has been violated? As Mill expresses it, does the occurrence of error mean that the syllogism is a bad one? Or can the error be accounted for otherwise? Is it possible that a process that would follow the rules of logic if it were spelled out is discernible even when the reasoning results in error?

---

[3] "A syllogism that is incompletely stated, in which one of the premises or the conclusion is tacitly present but not expressed, is called an enthymeme" (Cohen & Nagel, 1934, p. 78).

The distinction being made here is a familiar one in the psychology of learning and thinking. Thus Koffka (1935, Ch. 12) distinguishes between learning as accomplishment and the learning processes responsible for this accomplishment. Köhler's (1927) distinction between "good errors" and stupid ones is likewise relevant. Good errors, he points out, "may, in a certain sense, be absolutely appropriate to the situation" (p. 217), although they solve the problem no more than do stupid ones. Wertheimer (1959), too, distinguishes between solutions obtained by "blind" procedures and "fine, genuine solutions." Again the difference is one of process, since in both cases the result may be the same. In the same way, in connection with the present problem, we may ask: Given contrasting results—correct solutions and errors in deductive reasoning—what can we say about the thinking processes that account for them? Are the processes necessarily different because their effects are different?

Illustrative data that bear on this issue will be presented. They were obtained from 46 graduate students of psychology who were asked to evaluate the logical adequacy of deductions presented in the context of everyday problems. Most of the subjects had no training in formal logic.[4] The material was presented under group conditions, the subjects writing out their judgments and their grounds for making them. Instructions included an explicit statement that the logical adequacy of the arguments was to be judged, not the truth of the statements.

Individual interviews were also conducted with a different set of subjects. Although the results agree generally with those obtained under group conditions, they will not be presented because of a problem already mentioned. Mill, it has been seen, pointed to the difficulty of recognizing a bad syllogism because a person can easily introduce into his incompletely stated argument a new premise that will make the inference valid. It was often difficult to decide whether the material elicited by direct questioning consisted of new premises not entering into the original inference, or whether the subject was, as intended, making clear his original understanding of the premises. Although the group data are less rich than the individual interviews, and although the products of written communication are likely to be less spontaneous reflections of the thinking process than interviews, they exclude this possibility of overlooking true fallacies; indeed, if anything, they overestimate the incidence of bad syllogisms, since the reasoning is often incompletely described in the case of apparent errors.[5]

The present data will be used only to illustrate the reasoning processes in cases of error; no quantitative results will be presented. As many authors have shown, the incidence of error in deductive reasoning depends on the form of the syllogism and its contents, as well as on instructions to the subjects. Quantitative results would have relevance only to the particular conditions studied here, whereas an inquiry into the nature of the errors obtained might be of more general interest.

---

[4] It must be recognized, of course, that the lack of formal training in logic is insufficient guarantee of the naiveté of subjects. It can only be pleaded that we have no better criterion at the present time. Whatever the informal self-education of subjects in this respect, the results to be presented cannot be regarded as the product of formal instruction in logic.

[5] A sample interview may be found in Henle (1958).

Several processes may be distinguished which led to error in dealing with the presented material:

FAILURE TO ACCEPT THE LOGICAL TASK

More specifically, this means failure to distinguish between a conclusion that is logically valid and one that is factually correct or one with which the subject agrees. This source of error has already been reported by Henle and Michael (1956, p. 124).

A sample syllogism follows, along with responses in which errors occur because of failure to grasp or accept the logical task.

Syllogism 6. A group of women were discussing their household problems. Mrs. Shivers broke the ice by saying: "I'm so glad we're talking about these problems. It's so important to talk about things that are in our minds. We spend so much of our time in the kitchen that of course household problems are in our minds. So it is important to talk about them." (Does it follow that it is important to talk about them? Give your reasoning.)

Responses: "The conclusion does not follow. The women must talk about household problems because it is important to talk about their problems, not because the problem is in their minds."

"No. It is not important to talk about things that are in our minds unless they worry us, which is not the case."

"No. Just because one spends 'so much time' in the kitchen it does not necessarily follow that household problems are 'in our minds'."

It should be noted that subjects who failed to accept the logical task frequently gave correct responses that are just as irrelevant to the question of the relation of logic to the thinking process as are the errors just cited. A few examples follow:

"Yes. It could be very important to the individual doing the talking and possibly to some of those listening, because it is important for people to 'get a load off their chest,' but not for any other reason, unless in the process one or the other learns something new and of value."

"Yes. It seems obvious that problems which are in the forefront of one's mind bring more consideration to them and possibly newer aspects when they are discussed with another. Two heads may be better than one."

"Yes it does. By talking household problems, a problem can be solved or worked through."

The errors illustrated here clearly do not demonstrate an inability of the subjects to reason logically, since they have not accepted the logical task. They have evaluated the content of the conclusion, not the logical form of the argument. Richter (1957, p. 341) makes the same distinction, carrying the analysis a step farther. Apart from careless mistakes and those arising from "imperfections in the classifying operation"—i.e., from an inability to make logical deductions, he describes errors arising from "a general failure to grasp the concept of 'logical validity'" and those arising from "the specific inability to differentiate 'logical validity' from another attribute of syllogisms," namely their factual status. To apply to the present data, the factual criterion needs to be interpreted to mean what is believed to be true or reasonable as well as what is known to be true. With Richter's distinction in mind, we may note that no subject failed completely to grasp the logical task. Still, one individual judged on the basis of belief rather than logical validity in the case of 9 out of 10 syllogisms, and several others judged in this manner with at least half of the syllogisms presented.

RESTATEMENT OF A PREMISE OR CONCLUSION SO THAT THE INTENDED MEANING IS CHANGED

In a number of cases examination of the material from which a subject reasoned showed that it differed from that which was originally presented. In such cases the validity of the argument can, of course, be judged only in relation to the syllogism actually employed, not the one intended.

1. In the case of syllogisms in which a conclusion from two particular premises was to be evaluated, the premises were occasionally restated as universals, from which the conclusion followed. The following syllogism permitted this kind of change:

Syllogism 8. Mrs. Cooke had studied home economics in college. "Youth is a time of rapid growth and great demands on energy," she said. "Many youngsters don't get enough vitamins in their daily diet. And since some vitamin deficiencies are dangerous to health, it follows that the health of many of our youngsters is being endangered by inadequate diet." (Does it follow that the health of many youngsters is being endangered by inadequate diet? Give your reasoning.)

Responses: "Yes. Youngsters (A) don't get vitamins (B). Not getting vitamins (B) dangerous to health (C). Youngsters (A) in danger of poor health (C)."

"This follows. . . . Youth doesn't get enough vitamins. Vitamins are necessary or bad health results. Therefore youth are endangered. All of these follow logically, and I believe a correct inference has been made."

2. Other changes were introduced into the premises by subjects whose reasoning was found to be correct even though their answers to the intended problem were wrong. The same syllogism (Syllogism 8) will be used to illustrate.

Responses: "There is some question about the equivalence of 'don't get enough vitamins' and a vitamin deficiency dangerous to health. If one assumes their equivalence, then the conclusion just about manages to follow."

"Correct if we assume that the youngsters are lacking those vitamins in their diet which endanger health."

"It seems to follow, assuming that the deficient vitamins are also the vital ones."

Everyday discussions permit greater scope for such changes in meaning than does the material employed here. Private meanings of terms or idiosyncratic or unconscious equations of concepts may enter in to prevent two people from drawing the same conclusion even though they start ostensibly from the same material.

OMISSION OF A PREMISE

Occasionally a subject employed only one of the presented premises. The informal manner in which the premises were set out made possible this treatment of the material. Syllogism 6 (cf. above) will be used to illustrate.

Responses: "Not correct logically. (*a*) A group of women spend much of their time in the kitchen; (*b*) thus household problems occupy their minds. . . . (*c*) Therefore it is important to talk about them. Mrs. Shivers' subjective thinking."

"Doesn't follow from the preceding statement. Because she spends a lot of time in the kitchen and household problems are on her mind, it doesn't follow that it is important to talk about them."

In a few cases, indeed, both of the intended premises were omitted by the subject. For example (Syllogism 6):

"No it doesn't follow. The fact that she spends much time in the kitchen has nothing to do with whether or not it is important to talk about the problems."

It will be seen that in all these cases in which a premise has been omitted, the subject correctly reports that the conclusion does not follow. If we disregard the intended premises and consider only the material actually used in reasoning, the subjects' deductions are seen to be correct.

SLIPPING IN OF ADDITIONAL PREMISES

Under the present conditions this device was infrequent, but it seems to be more common where premises are stated less completely, and where the issues are of more immediate practical concern, as in everyday discussion. For example, premises may be added that are so commonplace, so much taken for granted, that they escape attention.

Instances of the adding of assumptions to our material follow. Since they are readily understandable, and for the sake of brevity, the responses will be quoted without the presented syllogism.

Responses: "Without intending sophistry it must be pointed out that the implicit premise is 'crime is bad.' If this is granted, the conclusion follows that if comics contribute to influence the youth along criminal paths, they should be eliminated."

". . . If comic books are an evil influence, then they should be got rid of." Here the subject is clearly employing an implicit premise that he added to the material presented, namely "Whatever is an evil influence should be got rid of."

Again in these cases, wrong answers are obtained by correct reasoning if we consider the syllogism as the subject understood it rather than the one the investigator hoped to present.

When all the above processes have been taken into account, a considerable number of errors remain. Do these represent true fallacies? It is difficult to say since in these cases either no account of the reasoning process is supplied, or an unclear explanation is given, or else the subject merely sets out the premises and conclusion given to him without comment. It might be that many of these errors would have been found on examination to involve processes similar to the ones described above, so that no error in reasoning is involved. On the other hand, the possibility of true fallacies cannot be excluded.

In connection with the unexplained errors, it is worth noting that almost as many correct responses as errors were unexplained or unclearly explained. The absence of explanation is thus no reason to suspect a reasoning process of being fallacious.

It must also be mentioned that the processes described above as producing apparent errors although the deductive process is correct are by no means viewed as constituting a complete list. They were observed with particular materials and under particular conditions; and it is to be anticipated that further research will expand this list. Chapman and Chapman (1959), for example, have suggested several additional processes by which error is produced in syllogistic reasoning, one of which involves a misunderstanding of the task different from that described here.

Despite the limitations of the present data, they do show clearly that when subjects arrive at apparently invalid conclusions, or when they fail to spot a fallacy, they often do so because they have worked with materials different from those intended or because they have undertaken a task different from the one intended. In such cases, if we consider the materials and task as they were actually understood by individual subjects, we fail to find evidence of faulty reasoning. It must be concluded that the presence of error does not constitute evidence that the laws of logic are irrelevant to actual thinking. The data tend, rather, to support the older conception that these laws are widely discernible in the thinking process.

DISCUSSION

An examination of errors in syllogistic reasoning leads to two conclusions: (*a*) While the possibility of fallacy often cannot be excluded it is, in Mill's words, "scarcely ever possible decidedly to affirm that any argument involves a bad syllogism." (*b*) Where error occurs, it need not involve faulty reasoning, but may be a function of the individual's understanding of the task or the materials presented to him. Spinoza (1930) has stated our present conclusion in more general terms:

> When men make errors in calculation, the numbers which are in their minds are not those which are upon the paper. As far as their mind is concerned there is no error, although it seems as if there were, because we think that the numbers in their minds are those which are upon the paper. If we did not think so, we should not believe them to be in error. For example, when I lately heard a man complaining that his court had flown into one of his neighbour's fowls, I understood what he meant, and therefore did not imagine him to be in error. This is the source from which so many controversies arise—that men either do not properly explain their own thoughts, or do not properly interpret those of other people; for, in truth, when they most contradict one another, they either think the same things or something different, so that those things which they suppose to be errors and absurdities in another person are not so (from the *Ethics,* Scholium Proposition 47, Part 2).

Such a conclusion, which the present analysis supports, is of some importance if, as Mill (1874) writes, "to draw inferences has been said to be the great business of life" (p. 21). Our data suggest that the intelligent adult is not so inept at carrying on this business as several recent accounts have made him out to be.

In considering the place of logical inference in everyday life, it is necessary also to take into account what Aristotle (1945) has termed the "practical syllogism," the case in which

> the conclusion drawn from the two premisses becomes the action. For example, when you conceive that every man ought to walk and you yourself are a man, you immediately walk; or if you conceive that on a particular occasion no man ought to walk, and you yourself are a man, you immediately remain at rest. In both instances action follows unless there is some hindrance or compulsion. . . . Again, I need a covering, and a cloak is a covering, I need a cloak. What I need I ought to make; I need a cloak, I ought to make a cloak. And the conclusion "I ought to make a cloak" is an action. The action results from the beginning of the train of thought. If there is to be a cloak, such and such a thing is necessary, if this thing then something else; and one immediately acts accordingly. That the action is the conclusion is quite clear. . . . (701a).

Aristotle points out that this whole process need not be explicit.

> The mind does not stop and consider at all one of the two premisses, namely, the obvious one; for example, if walking is good for a man, one does not waste time over the premiss "I am myself a man." Hence such things as we do without calculation, we do quickly (701a).

Again, "My appetite says, I must drink; this is drink, says sensation or imagination or thought, and one immediately drinks."

If we include, as it seems we must, the practical syllogism in the "great business of life," that of drawing inferences, it becomes difficult indeed to accept the conclusion that human beings are unable to reason logically. A couple of contemporary examples will suggest how ubiquitous is the practical syllogism in everyday life. "It is too far to walk to the Public Library; I must take a subway or bus. The Fifth Avenue bus passes the Public Library. I must take the Fifth Avenue bus." "I do not want to wear the same dress two days in succession. I wore this dress yesterday; so I do not want to wear this dress today." It is difficult to see how the individual could cope with the ordinary tasks of life if the practical syllogism embodied techniques which are not, as one author quoted above put it, "the common property of the unsophisticated subject," if, indeed, it were not a natural mode of functioning of the conscious mind.

Furthermore, if people were unable to reason logically, so that each arrived at different conclusions from the same premises, it is difficult to see how they could understand each other, follow one another's thinking, reach common decisions, and work together.

It might be argued that in the case of the practical syllogism we are dealing not with an implicit logical process, but with a learning process in which each response is cued off by the preceding one. There seem to be at least two grounds for rejecting such an interpretation. (*a*) There are cases in which the practical syllogism leads to a solution which is genuinely novel for the individual. This is surely so for some of the examples cited above, at least the first time they occur. Indeed, if the responses were habitual ones, the reasoning process would not be necessary. (*b*) Our data obtained from conventional syllogisms suggest that a logical process often occurs even in cases in which the reasoning seems fallacious. Such data can thus not be interpreted in terms of the "learned drive" postulated by Dollard and Miller (1950, p. 120) "to make . . . explanations and plans seem logical." And since they sometimes lead to novel conclusions (cf. Wertheimer, 1938), they cannot be interpreted as habitual responses. In the interests of parsimony, therefore, a common explanation for the practical syllogism and the verbal one seems preferable. An interpretation in terms of an implicit logical process would fit both kinds of case.

Two further implications of the present data may be mentioned. They suggest an approach to the problem of the influence of needs and attitudes on the reasoning process. For the problem remains: why do motivational influences so often appear to impair thinking? It is a plausible hypothesis that these influences do not distort the reasoning process, as has frequently been stated or implied—indeed that they do not act at all on the reasoning process—but rather that they affect the materials with which thinking works. Although this is not the place for a detailed analysis, one or two specific effects of this kind may be suggested. (*a*) It may be that a strong attitude toward, or emotional involvement with, particular material is in part responsible for the difficulty which many unsophisticated subjects experience in distinguishing between drawing a conclusion that is logically valid and one that is believed to be correct. Over a wide range of practical situations these two tasks are not in conflict; and the distinction is thus not an easy one. There is evidence that attitudes and emotions may limit the ability to make distinctions (cf. Henle, 1955). If this suggestion is correct, the more personally relevant the material employed, the more difficult it will be to accept the logical task. Such an effect might go a long way toward explaining the report of a number of writers that individuals are unable to reason logically about emotionally toned material. (*b*) An attitude can select from among the possibilities that the material presents, singling out, for example, one among several possible meanings. It may be that such a process is responsible for the changes in the meaning of premises observed in our data. (For possibilities of further analysis of this kind, cf. Henle, 1955.)

That motivational processes may act on reasoning in the manner suggested here receives support from a recent study by Kopp (1960), who subjected to a similar analysis reasoning processes related to the delusions of paranoid schizophrenics. His subjects, like the present ones, made errors which were found to lie more in the premises from which the reasoning proceeded than in the actual drawing of inferences.

In this connection, it would be interesting to analyze Arieti's (1955) examples of "paleological" thinking in the terms of the present study. Unfortunately, the published data are too scant to permit such an analysis. The examples presented would be fallacies in terms of Aristotelian logic if one could assume that the premises as understood by the subject are the same as those understood by the examiner. That this is the case may be doubted in a number of instances; and the search for an archaic logic thus seems premature. Some examples will be given which seem to call for an analysis different from Arieti's.

Arieti cites the case of a woman who "needed to identify herself with the Virgin Mary because of the extreme closeness and spiritual kinship she felt for the Virgin Mary" (1955, p. 195). The reasoning process is described as follows: "The Virgin Mary was a virgin; I am a virgin; therefore, I am the Virgin Mary." In the absence of additional material, we may only speculate that the proposition "I am the Virgin Mary" was not a conclusion deduced from the stated premises, but that it belongs rather in the category of revealed truth or intuition, not in that of deductive inference. It seems likely to the present writer that the syllogism quoted above was in the nature of a justification of this intuition to the examiner rather than the grounds for belief in it. The argument itself could be evaluated only in the light of additional information.

Another kind of example of Arieti's (1955) illustrates what he considers to be the propensity of young children "to indulge in paleologic thinking":

> A girl, three years and nine months old, saw two nuns walking together, and told her mother, "Mommy, look at the twins!" She thought that the nuns were twins because they were dressed alike. The characteristic of being dressed alike, which twins often have, led to the identification with the nuns (p. 199).

Again a different interpretation suggests itself: rather than viewing it as a deduction, we may take this process more simply as an instance of an inaccurate concept. Arieti also discusses dream symbols as examples of paleologic thinking. It is particularly difficult for the present writer to view such symbols as inferences, as if they were the conclusions of processes of deductive reasoning. They seem, rather, to involve a much more immediate grasping of similarities.

It is amusing to note that Arieti (1955) regards some delusions of identification as instances of paleologic, and describes the "formal mechanism" by which they are produced in a way that could easily be restated as a valid Aristotelian syllogism:

> If A may be identified with B because they have a common quality, it will be sufficient for me to acquire a quality of the person I want to be identified with, in order to become that person (p. 207).

Further discussion of the problems raised by Arieti must await more complete data.

The present study has implications also for another problem. To label an inference fallacious, as is clear from the analysis of this paper, is to make a statement about the results of reasoning, but not about the process that is responsible for these results. It might be a matter of considerable interest to study systematically and in detail the reasoning involved in various fallacies. It might be found that many fallacies are produced, not by faulty reasoning, but by specific changes in the material as the subject understands it. Some of the syllogisms employed here involved the fallacy of the undistributed middle; here changes of premises from particulars to universals and other changes in the meaning of premises were not uncommon. Two of the arguments presented contained illicit conversions, but no relevant data were obtained since the subjects systematically ignored the conversion and centered their attention on other parts of the material.[6] In connection with illicit conversions, the discussion of Chapman and Chapman (1959) is of interest; these authors suggest implicit assumptions and changes in meaning that may be involved in the acceptance of such conversions. Mill's (1874) observation that fallacies frequently involve changing the premises has been cited above. He analyzes several examples of the fallacy *a dicto secundum quid ad dictum simpliciter*[7] in these terms (p. 561–562). Fallacies of ambiguity seem likewise to lend themselves to the kind of analysis suggested here.[8]

In conclusion, observations have been presented which are consistent with the view that the rules of the syllogism describe processes that the mind follows in deductive reasoning, even when the syllogistic form is not explicitly employed.[9] A word of caution is necessary at this point. The present observations apply to deductive reasoning only. Even within this category, a very limited number of logical forms has been studied; and the study has been concerned only with the evaluation of presented arguments, not with the construction of arguments of one's own. The investigation needs to be extended before the generality of the conclusion can be assessed. It must also be pointed out that there are many types of thinking to which logical analysis does not apply at all, for example many aspects of free association, fantasy, many creative processes. To apply logical analysis to such processes (for example, Arieti's discussion of dream symbols) is as much an error as to ignore the underlying logical structure of incompletely stated arguments.

---

[6] This unexpected finding cannot be attributed to an insensitivity of the subjects to logical processes. Rather the fault lies with the investigator who, in the effort to disguise the formal structure of the syllogism, included irrelevancies that were taken by many subjects as the material for evaluation.

[7] "This is committed when, in the premises, a proposition is asserted with a qualification, and the qualification lost sight of in the conclusion; or oftener, when a limitation or condition, though not asserted, is necessary to the truth of the proposition, but is forgotten when that proposition comes to be employed as a premise" (Mill, 1874, p. 561).

[8] It is interesting to note that Mill's (1874) view comes close to the hypothesis offered here. It is true that he describes fallacies of ratiocination as:

> those which have their seat in the ratiocinative or deductive part of the investigation of truth (p. 559).

However he adds that it is by the rules of the syllogism

> that a reasoner is compelled distinctly to make his election what premises he is prepared to maintain. The election made, there is generally so little difficulty in seeing whether the conclusion follows from the premises set out, that we might without much logical impropriety have merged this fourth class of fallacies in the fifth, or Fallacies of Confusion (p. 560).

The latter class includes:

> all those fallacies in which the source of error is not so much a false estimate of the probative force of known evidence, as an indistinct, indefinite, and fluctuating conception of what the evidence is (p. 563).

(This class includes errors involving ambiguity of terms, begging the question, and fallacies of irrelevance.)

SUMMARY

Two views of the relation of logic to thinking have been presented: the older view that logic is the science of thought, and the position more widely held at present that its relation to thinking is normative only. It has been acknowledged that ordinary thinking, more often than not, does not assume the form of the syllogism, and that errors in reasoning are frequent. But does error mean that the logical process has been violated? or can we account for it in other ways? Evidence has been presented that suggests that even where the thinking process results in error, it can often be shown that it does not violate the rules of the syllogism. Many errors were found to be accounted for not in terms of a breakdown of the deductive process itself, but rather in terms of changes in the material from which the reasoning proceeds. The problem of the role of logical processes in everyday life was discussed more generally, and some implications of this study for other problems were considered.

---

[9] If it could be established generally that the laws of logic are those of the human mind, this would not be to give the former the status of "mere facts" or to make them relative to particular conditions; rather it would be to recognize the requiredness in thinking (cf. Köhler, 1938, especially pp. 45, 52).

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

ARISTOTLE. *Movement of animals.* (Trans. by E. S. Forster) Cambridge: Harvard Univer. Press, 1945.

BOOLE, G. *An investigation of the laws of thought.* London: MacMillan, 1854.

BRUNER, J. S., GOODNOW, JACQUELINE J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

CHAPMAN, L. J., & CHAPMAN, JEAN P. Atmosphere effect re-examined. *J. exp. Psychol.*, 1959, **58**, 220–226.

COHEN, M. R. *A preface to logic.* New York: Holt, 1944.

COHEN, M. R., & NAGEL, E. *An introduction to logic and scientific method.* New York: Harcourt, Brace, 1934.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

HENLE, MARY. Some effects of motivational processes on cognition. *Psychol. Rev.*, 1955, **62**, 423–432.

HENLE, MARY. The birth and death of ideas. Paper read at Colorado Symposium on Creative Thinking, Boulder, 1958.

HENLE, MARY, & MICHAEL, MIRIAM. The influence of attitudes on syllogistic reasoning. *J. soc. Psychol.*, 1956, **44**, 115–127.

KANT, I. *Introduction to logic and essay on the mistaken subtilty of the four figures.* (Trans. by T. K. Abbott) London: Longman's, Green, 1885.

KEYNES, J. N. *Studies and exercises in formal logic.* (2nd ed.) London: MacMillan, 1887.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

KÖHLER, W. *The mentality of apes.* (Trans. by Ella Winter) (2nd ed.) New York: Harcourt, Brace, 1927.

KÖHLER, W. *The place of value in a world of facts.* New York: Liveright, 1938.

KOPP, S. Deductive reasoning in paranoid schizophrenics. Unpublished doctoral dissertation, New School for Social Research, 1960.

LEFFORD, A. The influence of emotional subject matter on logical reasoning. *J. gen. Psychol.,* 1946, **34**, 127–151.

MILL, J. S. *A system of logic.* (8th ed.) New York: Harper, 1874.

MORGAN, J. J. B., & MORTON, J. T. The distortion of syllogistic reasoning produced by personal convictions. *J. soc. Psychol.,* 1944, **20**, 39–59.

NAGEL, E. *Logic without metaphysics.* Glencoe, Ill.: Free Press, 1956.

RICHTER, M. N., JR. The theoretical interpretation of errors in syllogistic reasoning. *J. Psychol.,* 1957, **43**, 341–344.

RUSSELL, B. The axiom of infinity. *Hibbert J.,* 1904, **2**, 809–812.

SCHILLER, F. C. S. *Logic for use.* New York: Harcourt, Brace, 1930.

SPINOZA, B. *Selections.* (Ed. by J. Wild) New York: Scribner, 1930.

VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944. Pp. 104–113.

WERTHEIMER, M. The syllogism and productive thinking. In W. D. Ellis (Ed.), *A source book of gestalt psychology.* London: Kegan Paul, Trench, Trubner, 1938. Selection 23.

WERTHEIMER, M. *Productive thinkin* (Enlarged ed.) New York: Harp 1959.

(Received December 13, 1960)

ERRATA

In an article by Joan E. Jones, "All-or-None versus Incremental Learning," published in *Psychological Review,* 1962, **69**, 156–160, errors appeared in the numbering of experiments. On page 159 in Table 1, and in the last paragraph of the text on page 159 where Table 1 is discussed, where Experiment III appears it should be read as Experiment I.

In the table of contents on the front cover of the May 1962 issue, tiguity and Reinforcement in Relation to CS-UCS Intervals in Classical Aversive Conditioning."

VOL. 69, NO. 5 SEPTEMBER 1962

PSYCHOLOGICAL REVIEW

COGNITIVE, SOCIAL, AND PHYSIOLOGICAL DETERMINANTS OF EMOTIONAL STATE[1]

STANLEY SCHACHTER AND JEROME E. SINGER
Columbia University *Pennsylvania State University*

The problem of which cues, internal or external, permit a person to label and identify his own emotional state has been with us since the days that James (1890) first tendered his doctrine that "the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur *is* the emotion" (p. 449). Since we are aware of a variety of feeling and emotion states, it should follow from James' proposition that the various emotions will be accompanied by a variety of differentiable bodily states. Following James' pronouncement, a formidable number of studies were undertaken in search of the physiological differentiators of the emotions. The results, in these early days, were almost uniformly negative. All of the emotional states experimentally manipulated were characterized by a general pattern of excitation of the sympathetic nervous system but there appeared to be no clear-cut physiological discriminators of the various emotions. This pattern of results was so consistent from experiment to experiment that Cannon (1929) offered, as one of the crucial criticisms of the James-Lange theory, the fact that "the same visceral changes occur in very different emotional states and in non-emotional states" (p. 351).

More recent work, however, has given some indication that there may be differentiators. Ax (1953) and Schachter (1957) studied fear and anger. On a large number of indices both of these states were characterized by a similarly high level of autonomic activation but on several indices they did differ in the degree of activation. Wolf and Wolff (1947) studied a subject with a gastric fistula and were able to distinguish two patterns in the physiological responses of the stomach wall. It should be noted, though, that for many months they studied their subject during and following a great variety of moods and emotions and were able to distinguish only two patterns.

Whether or not there are physiological distinctions among the various emotional states must be considered an open question. Recent work might be taken to indicate that such differences are at best rather subtle and that the variety of emotion, mood, and feeling states are by no means matched by an equal variety of visceral patterns.

---

[1] This experiment is part of a program of research on cognitive and physiological determinants of emotional state which is being conducted at the Department of Social Psychology at Columbia University under PHS Research Grant M-2584 from the National Institute of Mental Health, United States Public Health Service. This experiment was conducted at the Laboratory for Research in Social Relations at the University of Minnesota.

The authors wish to thank Jean Carlin and Ruth Hase, the physicians in the study, and Bibb Latané and Leonard Weller who were the paid participants.

This rather ambiguous situation has led Ruckmick (1936), Hunt, Cole, and Reis (1958), Schachter (1959) and others to suggest that cognitive factors may be major determinants of emotional states. Granted a general pattern of sympathetic excitation as characteristic of emotional states, granted that there may be some differences in pattern from state to state, it is suggested that one labels, interprets, and identifies this stirred-up state in terms of the characteristics of the precipitating situation and one's apperceptive mass. This suggests, then, that an emotional state may be considered a function of a state of physiological arousal[2] and of a cognition appropriate to this state of arousal. The cognition, in a sense, exerts a steering function. Cognitions arising from the immediate situation as interpreted by past experience provide the framework within which one understands and labels his feelings. It is the cognition which determines whether the state of physiological arousal will be labeled as "anger," "joy," "fear," or whatever.

In order to examine the implications of this formulation let us consider the fashion in which these two elements, a state of physiological arousal and cognitive factors, would interact in a variety of situations. In most emotion inducing situations, of course, the two factors are completely interrelated. Imagine a man walking alone down a dark alley, a figure with a gun suddenly appears. The perception-cognition "figure with a gun" in some fashion initiates a state of physiological arousal; this state of arousal is interpreted in terms of knowledge about dark alleys and guns and the state of arousal is labeled "fear." Similarly a student who unexpectedly learns that he has made Phi Beta Kappa may experience a state of arousal which he will label "joy."

Let us now consider circumstances in which these two elements, the physiological and the cognitive, are, to some extent, independent. First, is the state of physiological arousal alone sufficient to induce an emotion? Best evidence indicates that it is not. Marañon[3] (1924), in a fascinating study, (which was replicated by Cantril & Hunt, 1932, and Landis & Hunt, 1932) injected 210 of his patients with the sympathomimetic agent adrenalin and then simply asked them to introspect. Seventy-one percent of his subjects simply reported their physical symptoms with no emotional overtones; 29% of the subjects responded in an apparently emotional fashion. Of these the great majority described their feelings in a fashion that Marañon labeled "cold" or "as if" emotions, that is, they made statements such as "I feel *as if* I were afraid" or "*as if* I were awaiting a great happiness." This is a sort of emotional "déjà vu" experience; these subjects are neither happy nor afraid, they feel "as if" they were. Finally a very few cases apparently reported a genuine emotional experience. However, in order to produce this reaction in most of these few cases, Marañon (1924) points out:

---

[2] Though our experiments are concerned exclusively with the physiological changes produced by the injection of adrenalin, which appear to be primarily the result of sympathetic excitation, the term physiological arousal is used in preference to the more specific "excitation of the sympathetic nervous system" because there are indications, to be discussed later, that this formulation is applicable to a variety of bodily states.

[3] Translated copies of Marañon's (1924) paper may be obtained by writing to the senior author.

> One must suggest a memory with strong affective force but not so strong as to produce an emotion in the normal state. For example, in several cases we spoke to our patients before the injection of their sick children or dead parents and they responded calmly to this topic. The same topic presented later, during the adrenal commotion, was sufficient to trigger emotion. This adrenal commotion places the subject in a situation of 'affective imminence' (pp. 307–308).

Apparently, then, to produce a genuinely emotional reaction to adrenalin, Marañon was forced to provide such subjects with an appropriate cognition.

Though Marañon (1924) is not explicit on his procedure, it is clear that his subjects knew that they were receiving an injection and in all likelihood knew that they were receiving adrenalin and probably had some order of familiarity with its effects. In short, though they underwent the pattern of sympathetic discharge common to strong emotional states, at the same time they had a completely appropriate cognition or explanation as to why they felt this way. This, we would suggest, is the reason so few of Marañon's subjects reported any emotional experience.

Consider now a person in a state of physiological arousal for which no immediately explanatory or appropriate cognitions are available. Such a state could result were one covertly to inject a subject with adrenalin or, unknown to him, feed the subject a sympathomimetic drug such as ephedrine. Under such conditions a subject would be aware of palpitations, tremor, face flushing, and most of the battery of symptoms associated with a discharge of the sympathetic nervous system. In contrast to Marañon's (1924) subjects he would, at the same time, be utterly unaware of why he felt this way. What would be the consequence of such a state?

Schachter (1959) has suggested that precisely such a state would lead to the arousal of "evaluative needs" (Festinger, 1954), that is, pressures would act on an individual in such a state to understand and label his bodily feelings. His bodily state grossly resembles the condition in which it has been at times of emotional excitement. How would he label his present feelings? It is suggested, of course, that he will label his feelings in terms of his knowledge of the immediate situation.[4] Should he at the time be with a beautiful woman he might decide that he was wildly in love or sexually excited. Should he be at a gay party, he might, by comparing himself to others, decide that he was extremely happy and euphoric. Should he be arguing with his wife, he might explode in fury and hatred. Or, should the situation be completely inappropriate he could decide that he was excited about something that had recently happened to him or, simply, that he was sick. In any case, it is our basic assumption that emotional states are a function of the interaction of such cognitive factors with a state of physiological arousal.

This line of thought, then, leads to the following propositions:

1. Given a state of physiological arousal for which an individual has no immediate explanation, he will "label" this state and describe his feelings in terms of the cognitions available to him. To the extent that cognitive factors are potent determiners of emotional states, it could be anticipated that precisely the same state of physiological arousal could be labeled "joy" or "fury" or "jealousy" or any of a great diversity

---

[4] This suggestion is not new for several psychologists have suggested that situational factors should be considered the chief differentiators of the emotions. Hunt, Cole, and Reis (1958) probably make this point most explicitly in their study distinguishing among fear, anger, and sorrow in terms of situational characteristics.

of emotional labels depending on the cognitive aspects of the situation.

2. Given a state of physiological arousal for which an individual has a completely appropriate explanation (e.g., "I feel this way because I have just received an injection of adrenalin") no evaluative needs will arise and the individual is unlikely to label his feelings in terms of the alternative cognitions available.

Finally, consider a condition in which emotion inducing cognitions are present but there is no state of physiological arousal. For example, an individual might be completely aware that he is in great danger but for some reason (drug or surgical) remain in a state of physiological quiescence. Does he experience the emotion "fear"? Our formulation of emotion as a joint function of a state of physiological arousal and an appropriate cognition, would, of course, suggest that he does not, which leads to our final proposition.

3. Given the same cognitive circumstances, the individual will react emotionally or describe his feelings as emotions only to the extent that he experiences a state of physiological arousal.[5]

PROCEDURE

The experimental test of these propositions requires (*a*) the experimental manipulation of a state of physiological arousal, (*b*) the manipulation of the extent to which the subject has an appropriate or proper explanation of his bodily state, and (*c*) the creation of situations from which explanatory cognitions may be derived.

In order to satisfy the first two experimental requirements, the experiment was cast in the framework of a study of the effects of vitamin supplements on vision. As soon as a subject arrived, he was taken to a private room and told by the experimenter:

> In this experiment we would like to make various tests of your vision. We are particularly interested in how certain vitamin compounds and vitamin supplements affect the visual skills. In particular, we want to find out how the vitamin compound called 'Suproxin' affects your vision.
>
> What we would like to do, then, if we can get your permission, is to give you a small injection of Suproxin. The injection itself is mild and harmless; however, since some people do object to being injected we don't want to talk you into anything. Would you mind receiving a Suproxin injection?

If the subject agrees to the injection (and all but 1 of 185 subjects did) the experimenter continues with instructions we shall describe shortly, then leaves the room. In a few minutes a physician enters the room, briefly repeats the experimenter's instructions, takes the subject's pulse and then injects him with Suproxin.

Depending upon condition, the subject receives one of two forms of Suproxin—epinephrine or a placebo.

Epinephrine or adrenalin is a sympathomimetic drug whose effects, with minor exceptions, are almost a perfect mimicry of a discharge of the sympathetic nervous system. Shortly after injection systolic blood pressure increases markedly, heart rate increases somewhat, cutaneous blood flow decreases, while muscle and cerebral blood flow increase, blood sugar and lactic acid concentration increase, and respiration rate increases slightly. As far as the subject is concerned the major subjective symptoms are palpitation, tremor, and sometimes a feeling of flushing and accelerated breathing. With a subcutaneous injection (in the dosage administered to our subjects), such effects usually begin within 3–5 minutes of injection and last anywhere from 10 minutes to an hour. For most subjects these effects are dissipated within 15–20 minutes after injection.

Subjects receiving epinephrine received a subcutaneous injection of ½ cubic centimeter of a 1 : 1000 solution of Winthrop Laboratory's Suprarenin, a saline solution of epinephrine bitartrate.

Subjects in the placebo condition received a subcutaneous injection of ½ cubic centimeter of saline solution. This is, of course, completely neutral material with no side effects at all.

---

[5] In his critique of the James-Lange theory of emotion, Cannon (1929) also makes the point that sympathectomized animals and patients do seem to manifest emotional behavior. This criticism is, of course, as applicable to the above proposition as it was to the James-Lange formulation. We shall discuss the issues involved in later papers.

Manipulating an Appropriate Explanation

By "appropriate" we refer to the extent to which the subject has an authoritative, unequivocal explanation of his bodily condition. Thus, a subject who had been informed by the physician that as a direct consequence of the injection he would feel palpitations, tremor, etc. would be considered to have a completely appropriate explanation. A subject who had been informed only that the injection would have no side effects would have no appropriate explanation of his state. This dimension of appropriateness was manipulated in three experimental conditions which shall be called: Epinephrine Informed (Epi Inf), Epinephrine Ignorant (Epi Ign), and Epinephrine Misinformed (Epi Mis).

Immediately after the subject had agreed to the injection and before the physician entered the room, the experimenter's spiel in each of these conditions went as follows:

> *Epinephrine Informed.* I should also tell you that some of our subjects have experienced side effects from the Suproxin. These side effects are transitory, that is, they will only last for about 15 or 20 minutes. What will probably happen is that your hand will start to shake, your heart will start to pound, and your face may get warm and flushed. Again these are side effects lasting about 15 or 20 minutes.

While the physician was giving the injection, she told the subject that the injection was mild and harmless and repeated this description of the symptoms that the subject could expect as a consequence of the shot. In this condition, then, subjects have a completely appropriate explanation of their bodily state. They know precisely what they will feel and why.

Epinephrine Ignorant. In this condition, when the subject agreed to the injection, the experimenter said nothing more relevant to side effects and simply left the room. While the physician was giving the injection, she told the subject that the injection was mild and harmless and would have no side effects. In this condition, then, the subject has no experimentally provided explanation for his bodily state.

> *Epinephrine Misinformed.* I should also tell you that some of our subjects have experienced side effects from the Suproxin. These side effects are transitory, that is, they will only last for about 15 or 20 minutes. What will probably happen is that your feet will feel numb, you will have an itching sensation over parts of your body, and you may get a slight headache. Again these are side effects lasting 15 or 20 minutes.

And again, the physician repeated these symptoms while injecting the subject.

None of these symptoms, of course, are consequences of an injection of epinephrine and, in effect, these instructions provide the subject with a completely inappropriate explanation of his bodily feelings. This condition was introduced as a control condition of sorts. It seemed possible that the description of side effects in the Epi Inf condition might turn the subject introspective, self-examining, possibly slightly troubled. Differences on the dependent variable between the Epi Inf and Epi Ign conditions might, then, be due to such factors rather than to differences in appropriateness. The false symptoms in the Epi Mis condition should similarly turn the subject introspective, etc., but the instructions in this condition do not provide an appropriate explanation of the subject's state.

Subjects in all of the above conditions were injected with epinephrine. Finally, there was a placebo condition in which subjects, who were injected with saline solution, were given precisely the same treatment as subjects in the Epi Ign condition.

Producing an Emotion Inducing Cognition

Our initial hypothesis has suggested that given a state of physiological arousal for which the individual has no adequate explanation, cognitive factors can lead the individual to describe his feelings with any of a diversity of emotional labels. In order to test this hypothesis, it was decided to manipulate emotional states which can be considered quite different—euphoria and anger.

There are, of course, many ways to induce such states. In our own program of research, we have concentrated on social determinants of emotional states and have been able to demonstrate in other studies that people do evaluate their own feelings by comparing themselves with others around them (Schachter 1959; Wrightsman 1960). In this experiment we have attempted again to manipulate emotional state by social means. In one set of conditions, the subject is placed together with a stooge who has been trained to act euphorically. In a second set of conditions the subject is with a stooge trained to act in an angry fashion.

Euphoria

Immediately[6] after the subject had been injected, the physician left the room and the experimenter returned with a stooge whom he introduced as another subject, then said:

> Both of you have had the Suproxin shot and you'll both be taking the same tests of vision. What I ask you to do now is just wait for 20 minutes. The reason for this is simply that we have to allow 20 minutes for the Suproxin to get from the injection site into the bloodstream. At the end of 20 minutes when we are certain that most of the Suproxin has been absorbed into the bloodstream, we'll begin the tests of vision.

The room in which this was said had been deliberately put into a state of mild disarray. As he was leaving, the experimenter apologetically added:

> The only other thing I should do is to apologize for the condition of the room. I just didn't have time to clean it up. So, if you need any scratch paper or rubber bands or pencils, help yourself. I'll be back in 20 minutes to begin the vision tests.

As soon as the experimenter had left, the stooge introduced himself again, made a series of standard icebreaker comments, and then launched his routine. For observation purposes, the stooge's act was broken into a series of standard units, demarcated by a change in activity or a standard comment. In sequence, the units of the stooge's routine were the following:

1. Stooge reaches for a piece of paper and starts doodling saying, "They said we could use this for scratch, didn't they?" He doodles a fish for some 30 seconds, then says:

2. "This scrap paper isn't even much good for doodling" and crumples paper and attempts to throw it into wastebasket in far corner of the room. He misses but this leads him into a "basketball game." He crumples up other sheets of paper, shoots a few baskets, says "Two points" occasionally. He gets up and does a jump shot saying, "The old jump shot is really on today."

3. If the subject has not joined in, the stooge throws a paper basketball to the subject saying, "Here, you try it."

4. Stooge continues his game saying, "The trouble with paper basketballs is that you don't really have any control."

5. Stooge continues basketball, then gives it up saying, "This is one of my good days. I feel like a kid again. I think I'll make a plane." He makes a paper airplane saying, "I guess I'll make one of the longer ones."

6. Stooge flies plane. Gets up and retrieves plane. Flies again, etc.

7. Stooge throws plane at subject.

8. Stooge, flying plane, says, "Even when I was a kid, I was never much good at this."

9. Stooge tears off part of plane saying, "Maybe this plane can't fly but at least it's good for something." He wads up paper and making a slingshot of a rubber band begins to shoot the paper.

10. Shooting, the stooge says, "They [paper ammunition] really go better if you make them long. They don't work right if you wad them up."

11. While shooting, stooge notices a sloppy pile of manila folders on a table. He builds a tower of these folders, then goes to the opposite end of the room to shoot at the tower.

12. He misses several times, then hits and cheers as the tower falls. He goes over to pick up the folders.

13. While picking up, he notices, behind a portable blackboard, a pair of hula hoops which have been covered with black tape with a few wires sticking out of the tape. He reaches for these, taking one for himself and putting the other aside but within reaching distance of the subject. The stooge tries the hula hoop, saying, "This isn't as easy as it looks."

14. Stooge twirls hoop wildly on arm, saying, "Hey, look at this—this is great."

15. Stooge replaces the hula hoop and sits down with his feet on the table. Shortly thereafter the experimenter returns to the room.

---

[6] It was, of course, imperative that the sequence with the stooge begin before the subject felt his first symptoms for otherwise the subject would be virtually forced to interpret his feelings in terms of events preceding the stooge's entrance. Pretests had indicated that, for most subjects, epinphrine-caused symptoms began within 3–5 minutes after injection. A deliberate attempt was made then to bring in the stooge within 1 minute after the subject's injection.

This routine was completely standard, though its pace, of course, varied depending upon the subject's reaction, the extent to which he entered into this bedlam and the extent to which he initiated activities of his own. The only variations from this standard routine were those forced by the subject. Should the subject originate some nonsense of his own and request the stooge to join in, he would do so. And, he would, of course, respond to any comments initiated by the subject.

Subjects in each of the three "appropriateness" conditions and in the placebo condition were submitted to this setup. The stooge, of course, never knew in which condition any particular subject fell.

Anger

Immediately after the injection, the experimenter brought a stooge into the subject's room, introduced the two and after explaining the necessity for a 20 minute delay for "the Suproxin to get from the injection site into the bloodstream" he continued, "We would like you to use these 20 minutes to answer these questionnaires." Then handing out the questionnaires, he concludes with, "I'll be back in 20 minutes to pick up the questionnaires and begin the tests of vision."

Before looking at the questionnaire, the stooge says to the subject,

> I really wanted to come for an experiment today, but I think it's unfair for them to give you shots. At least, they should have told us about the shots when they called us; you hate to refuse, once you're here already.

The questionnaires, five pages long, start off innocently requesting face sheet information and then grow increasingly personal and insulting. The stooge, sitting directly opposite the subject, paces his own answers so that at all times subject and stooge are working on the same question. At regular points in the questionnaire, the stooge makes a series of standardized comments about the questions. His comments start off innocently enough, grow increasingly querulous, and finally he ends up in a rage. In sequence, he makes the following comments.

1. Before answering any items, he leafs quickly through the questionnaire saying, "Boy, this is a long one."

2. Question 7 on the questionnaire requests, "List the foods that you would eat in a typical day." The stooge comments, "Oh for Pete's sake, what did I have for breakfast this morning?"

3. Question 9 asks, "Do you ever hear bells? ________. How often? ________." The stooge remarks, "Look at Question 9. How ridiculous can you get? I hear bells every time I change classes."

4. Question 13 requests, "List the childhood diseases you have had and the age at which you had them" to which the stooge remarks, "I get annoyed at this childhood disease question. I can't remember what childhood diseases I had, and especially at what age. Can you?"

5. Question 17 asks "What is your father's average annual income?" and the stooge says, "This really irritates me. It's none of their business what my father makes. I'm leaving that blank."

6. Question 25 presents a long series of items such as "Does not bathe or wash regularly," "Seems to need psychiatric care," etc. and requests the respondent to write down for which member of his immediate family each item seems most applicable. The question specifically prohibits the answer "None" and each item must be answered. The stooge says, "I'll be damned if I'll fill out Number 25. 'Does not bathe or wash regularly'—that's a real insult." He then angrily crosses out the entire item.

7. Question 28 reads:

"How many times each week do you have sexual intercourse?" 0–1 ____ 2–3 ____ 4–6 ____ 7 and over ____. The stooge bites out, "The hell with it! I don't have to tell them all this."

8. The stooge sits sullenly for a few moments then he rips up his questionnaire, crumples the pieces and hurls them to the floor, saying, "I'm not wasting any more time. I'm getting my books and leaving" and he stamps out of the room.

9. The questionnaire continues for eight more questions ending with: "With how many men (other than your father) has your mother had extramarital relationships?"

4 and under ____: 5–9 ____: 10 and over ____.

Subjects in the Epi Ign, Epi Inf and Placebo conditions were run through this "anger" inducing sequence. The stooge, again, did not know to which condition the subject had been assigned.

In summary, this is a seven condition experiment which, for two different emotional states, allows us (*a*) to evaluate the effects of "appropriateness" on emotional inducibility and (*b*) to begin to evaluate the effects of sympathetic activation on emotional inducibility. In schematic form the conditions are the following:

EUPHORIA	ANGER
Epi Inf	Epi Inf
Epi Ign	Epi Ign
Epi Mis	Placebo
Placebo	

The Epi Mis condition was not run in the Anger sequence. This was originally conceived as a control condition and it was felt that its inclusion in the Euphoria conditions alone would suffice as a means of evaluating the possible artifactual effect of the Epi Inf instructions.

Measurement

Two types of measures of emotional state were obtained. Standardized observation through a one-way mirror was the technique used to assess the subject's behavior. To what extent did he act euphoric or angry? Such behavior can be considered in a way as a "semiprivate" index of mood for as far as the subject was concerned, his emotional behavior could be known only to the other person in the room—presumably another student. The second type of measure was self-report in which, on a variety of scales, the subject indicated his mood of the moment. Such measures can be considered "public" indices of mood for they would, of course, be available to the experimenter and his associates.

Observation

Euphoria. For each of the first 14 units of the stooge's standardized routine an observer kept a running chronicle of what the subject did and said. For each unit the observer coded the subject's behavior in one or more of the following categories:

Category 1: Joins in activity. If the subject entered into the stooge's activities, e.g., if he made or flew airplanes, threw paper basketballs, hula hooped, etc., his behavior was coded in this category.

Category 2: Initiates new activity. A subject was so coded if he gave indications of creative euphoria, that is, if, on his own, he initiated behavior outside of the stooge's routine. Instances of such behavior would be the subject who threw open the window and, laughing, hurled paper basketballs at passersby; or, the subject who jumped on a table and spun one hula hoop on his leg and the other on his neck.

Categories 3 and 4: Ignores or watches stooge. Subjects who paid flatly no attention to the stooge or who, with or without comment, simply watched the stooge without joining in his activity were coded in these categories.

For any particular unit of behavior, the subject's behavior was coded in one or more of these categories. To test reliability of coding two observers independently coded two experimental sessions. The observers agreed completely on the coding of 88% of the units.

Anger. For each of the units of stooge behavior, an observer recorded the subject's responses and coded them according to the following category scheme:

Category 1: Agrees. In response to the stooge the subject makes a comment indicating that he agrees with the stooge's standardized comment or that he, too, is irked by a particular item on the questionnaire. For example, a subject who responded to the stooge's comment on the "father's income" question by saying, "I don't like that kind of personal question either" would be so coded (scored +2).

Category 2: Disagrees. In response to the stooge's comment, the subject makes a comment which indicates that he disagrees with the stooge's meaning or mood; e.g., in response to the stooge's comment on the "father's income" question, such a subject might say, "Take it easy, they probably have a good reason for wanting the information" (scored −2).

Category 3: Neutral. A noncommittal or irrelevant response to the stooge's remark (scored 0).

Category 4: Initiates agreement or disagreement. With no instigation by the stooge, a subject, so coded, would have volunteered a remark indicating that he felt the same way or, alternatively, quite differently than the stooge. Examples would be "Boy I hate this kind of thing" or "I'm enjoying this" (scored +2 or −2).

Category 5: Watches. The subject makes no verbal response to the stooge's comment but simply looks directly at him (scored 0).

Category 6: Ignores. The subject makes no verbal response to the stooge's comment nor does he look at him; the subject, paying no attention at all to the stooge, simply works at his own questionnaire (scored −1).

A subject was scored in one or more of these categories for each unit of stooge behavior. To test reliability, two observers independently coded three experimental sessions. In order to get a behavioral index of anger, observation protocol was scored according to the values presented in parentheses after each of the above definitions of categories. In a unit-by-unit comparison, the two observers agreed completely on the scoring of 71% of the units jointly observed. The scores of the two observers differed by a value of 1 or less for 88% of the units coded and in not a single case did the two observers differ in the direction of their scoring of a unit.

Self Report of Mood and Physical Condition

When the subject's session with the stooge was completed, the experimenter returned to the room, took pulses and said:

> Before we proceed with the vision tests, there is one other kind of information which we must have. We have found, as you can probably imagine, that there are many things beside Suproxin that affect how well you see in our tests. How hungry you are, how tired you are, and even the mood you're in at the time—whether you feel happy or irritated at the time of testing will affect how well you see. To understand the data we collect on you, then, we must be able to figure out which effects are due to causes such as these and which are caused by Suproxin.
>
> The only way we can get such information about your physical and emotional state is to have you tell us. I'll hand out these questionnaires and ask you to answer them as accurately as possible. Obviously, our data on the vision tests will only be as accurate as your description of your mental and physical state.

In keeping with this spiel, the questionnaire that the experimenter passed out contained a number of mock questions about hunger, fatigue, etc., as well as questions of more immediate relevance to the experiment. To measure mood or emotional state the following two were the crucial questions:

1. How irritated, angry or annoyed would you say you feel at present?

I don't feel at all irritated or angry (0)	I feel a little irritated and angry (1)	I feel quite irritated and angry (2)	I feel very irritated and angry (3)	I feel extremely irritated and angry (4)

2. How good or happy would you say you feel at present?

I don't feel at all happy or good (0)	I feel a little happy and good (1)	I feel quite happy and good (2)	I feel very happy and good (3)	I feel extremely happy and good (4)

To measure the physical effects of epinephrine and determine whether or not the injection had been successful in producing the necessary bodily state, the following questions were asked:

1. Have you experienced any palpitation (consciousness of your own heart beat)?

Not at all (0)	A slight amount (1)	A moderate amount (2)	An intense amount (3)

2. Did you feel any tremor (involuntary shaking of the hands, arms or legs)?

Not at all (0)	A slight amount (1)	A moderate amount (2)	An intense amount (3)

To measure possible effects of the instructions in the Epi Mis condition, the following questions were asked:

1. Did you feel any numbness in your feet?

2. Did you feel any itching sensation?

3. Did you experience any feeling of headache?

To all three of these questions was attached a four-point scale running from "Not at all" to "An intense amount."

In addition to these scales, the subjects were asked to answer two open-end questions on other physical or emotional sensations they may have experienced during the experimental session. A final measure of bodily state was pulse rate which was taken by the physician or the experimenter at two times—immediately before the injection and immediately after the session with the stooge.

When the subjects had completed these questionnaires, the experimenter announced that the experiment was over, explained the deception and its necessity in detail, answered any questions, and swore the subjects to secrecy. Finally, the subjects answered a brief questionnaire about their experiences, if any, with adrenalin and their previous knowledge or suspicion of the experimental setup. There was no indication that any of the subjects had known about the experiment beforehand but 11 subjects were so extremely suspicious of some crucial feature of the experiment that their data were automatically discarded.

Subjects

The subjects were all male, college students taking classes in introductory psychology at the University of Minnesota. Some 90% of the students in these classes volunteer for a subject pool for which they receive two extra points on their final exam for every hour that they serve as experimental subjects. For this study the records of all potential subjects were cleared with the Student Health Service in order to insure that no harmful effects would result from the injections.

Evaluation of the Experimental Design

The ideal test of our propositions would require circumstances which our experiment is far from realizing. First, the proposition that: "A state of physiological arousal for which an individual has no immediate explanation will lead him to label this state in terms of the cognitions available to him" obviously requires conditions under which the subject does not and cannot have a proper explanation of his bodily state. Though we toyed with such fantasies as ventilating the experimental room with vaporized adrenalin, reality forced us to rely on the disguised injection of Suproxin—a technique which was far from ideal for no matter what the experimenter told them, some subjects would inevitably attribute their feelings to the injection. To the extent that subjects did so, differences between the several appropriateness conditions should be attenuated.

Second, the proposition that: "Given the same cognitive circumstances the individual will react emotionally only to the extent that he experiences a state of physiological arousal" requires for its ideal test the manipulation of states of physiological arousal and of physiological quiescence. Though there is no question that epinephrine effectively produces a state of arousal, there is also no question that a placebo does not prevent physiological arousal. To the extent that the experimental situation effectively produces sympathetic stimulation in placebo subjects, the proposition is difficult to test, for such a factor would attenuate differences between epinephrine and placebo subjects.

Both of these factors, then, can be expected to interfere with the test of our several propositions. In presenting the results of this study, we shall first present condition by condition results and then evaluate the effect of these two factors on experimental differences.

RESULTS

Effects of the Injections on Bodily State

Let us examine first the success of the injections at producing the bodily state required to examine the propositions at test. Does the injection of epinephrine produce symptoms of sympathetic discharge as compared with the placebo injection? Relevant data are presented in Table 1 where it can be immediately seen that on all items subjects who were in epinephrine conditions show considerably more evidence of sympathetic activation than do subjects in placebo conditions. In all epinephrine conditions pulse rate increases significantly when compared with the decrease characteristic of the placebo conditions. On the scales it is clear that epinephrine subjects experi-

TABLE 1

THE EFFECTS OF THE INJECTIONS ON BODILY STATE

Condition	*N*	Pulse		Self-rating of				
		Pre	Post	Palpitation	Tremor	Numbness	Itching	Headache
Euphoria								
Epi Inf	27	85.7	88.6	1.20	1.43	0	0.16	0.32
Epi Ign	26	84.6	85.6	1.83	1.76	0.15	0	0.55
Epi Mis	26	82.9	86.0	1.27	2.00	0.06	0.08	0.23
Placebo	26	80.4	77.1	0.29	0.21	0.09	0	0.27
Anger								
Epi Inf	23	85.9	92.4	1.26	1.41	0.17	0	0.11
Epi Ign	23	85.0	96.8	1.44	1.78	0	0.06	0.21
Placebo	23	84.5	79.6	0.59	0.24	0.14	0.06	0.06

ence considerably more palpitation and tremor than do placebo subjects. In all possible comparisons on these symptoms, the mean scores of subjects in any of the epinephrine conditions are greater than the corresponding scores in the placebo conditions at better than the .001 level of significance. Examination of the absolute values of these scores makes it quite clear that subjects in epinephrine conditions were, indeed, in a state of physiological arousal, while most subjects in placebo conditions were in a relative state of physiological quiescence.

The epinephrine injection, of course, did not work with equal effectiveness for all subjects; indeed for a few subjects it did not work at all. Such subjects reported almost no palpitation or tremor, showed no increase in pulse and described no other relevant physical symptoms. Since for such subjects the necessary experimental conditions were not established, they were automatically excluded from the data and all further tabular presentations will not include such subjects. Table 1, however, does include the data of these subjects. There were four such subjects in euphoria conditions and one of them in anger conditions.

In order to evaluate further data on Epi Mis subjects it is necessary to note the results of the "numbness," "itching," and "headache" scales also presented in Table 1. Clearly the subjects in the Epi Mis condition do not differ on these scales from subjects in any of the other experimental conditions.

Effects of the Manipulations on Emotional State

Euphoria: Self-report. The effects of the several manipulations on emotional state in the euphoria conditions are presented in Table 2. The scores recorded in this table are derived, for each subject, by subtracting the value of the point he checks on the irritation scale from the value of the point he checks on the happiness scale. Thus, if a subject were to check the point "I feel a little irritated and angry" on the irritation scale and the point "I feel very happy and good" on the happiness scale, his score would be +2. The higher the positive value, the happier and better the subject reports himself as feeling. Though we employ an index for expositional simplicity, it should be noted that the two components of the index each yield results completely consistent with those obtained by use of this index.

Let us examine first the effects of the appropriateness instructions. Comparison of the scores for the Epi Mis and Epi Inf conditions makes it immediately clear that the experimental differences are not due to artifacts resulting from the informed instructions. In both conditions the subject was warned to expect a variety of symptoms as a consequence of the injection. In the Epi Mis condition, where the symptoms were inappropriate to the subject's bodily state the self-report score is almost twice that in the Epi Inf condition where the symptoms were completely appropriate to the subject's bodily state. It is reasonable, then, to attribute differences between informed subjects and those in other conditions to differences in manipulated appropriateness rather than to artifacts such as introspectiveness or self-examination.

It is clear that, consistent with expectations, subjects were more susceptible to the stooge's mood and consequently more euphoric when they had no explanation of their own bodily states than when they did. The means of both the Epi Ign and Epi Mis conditions are considerably greater than the mean of the Epi Inf condition.

It is of interest to note that Epi Mis subjects are somewhat more euphoric

TABLE 2

SELF-REPORT OF EMOTIONAL STATE IN THE EUPHORIA CONDITIONS

Condition	N	Self-Report scales	Comparison	p^a
Epi Inf	25	0.98	Epi Inf vs. Epi Mis	<.01
Epi Ign	25	1.78	Epi Inf vs. Epi Ign	.02
Epi Mis	25	1.90	Placebo vs. Epi Mis, Ign, or Inf	*ns*
Placebo	26	1.61		

All p values reported throughout paper are two-tailed.

than are Epi Ign subjects. This pattern repeats itself in other data shortly to be presented. We would attribute this difference to differences in the appropriateness dimension. Though, as in the Epi Ign condition, a subject is not provided with an explanation of his bodily state, it is, of course, possible that he will provide one for himself which is not derived from his interaction with the stooge. Most reasonably he could decide for himself that he feels this way because of the injection. To the extent that he does so he should be less susceptible to the stooge. It seems probable that he would be less likely to hit on such an explanation in the Epi Mis condition than in the Epi Ign condition for in the Epi Mis condition both the experimenter and the doctor have told him that the effects of the injection would be quite different from what he actually feels. The effect of such instructions is probably to make it more difficult for the subject himself to hit on the alternative explanation described above. There is some evidence to support this analysis. In open-end questions in which subjects described their own mood and state, 28% of the subjects in the Epi Ign condition made some connection between the injection and their bodily state compared with the 16% of subjects in the Epi Mis condition who did so. It could be considered, then, that these three conditions fall along a dimension of appropriateness, with the Epi Inf condition at one extreme and the Epi Mis condition at the other.

Comparing the placebo to the epinephrine conditions, we note a pattern which will repeat itself throughout the data. Placebo subjects are less euphoric than either Epi Mis or Epi Ign subjects but somewhat more euphoric than Epi Inf subjects. These differences are not, however, statistically significant. We shall consider the epinephrine-placebo comparisons in detail in a later section of this paper following the presentation of additional relevant data. For the moment, it is clear that, by self-report manipulating appropriateness has had a very strong effect on euphoria.

Behavior. Let us next examine the extent to which the subject's behavior was affected by the experimental manipulations. To the extent that his mood has been affected, one should expect that the subject will join in the stooge's whirl of manic activity and initiate similar activities of his own. The relevant data are presented in Table 3. The column labeled "Activity

TABLE 3

BEHAVIORAL INDICATIONS OF EMOTIONAL STATE IN THE EUPHORIA CONDITIONS

Condition	N	Activity index	Mean number of acts initiated
Epi Inf	25	12.72	.20
Epi Ign	25	18.28	.56
Epi Mis	25	22.56	.84
Placebo	26	16.00	.54

p value

Comparison	Activity index	Initiates
Epi Inf vs. Epi Mis	.05	.03
Epi Inf vs. Epi Ign	*ns*	.08
Plac vs. Epi Mis, Ign, or Inf	*ns*	*ns*

[a] Tested by X^2 comparison of the proportion of subjects in each condition initiating new acts.

index" presents summary figures on the extent to which the subject joined in the stooge's activity. This is a weighted index which reflects both the nature of the activities in which the subject engaged and the amount of time he was active. The index was devised by assigning the following weights to the subject's activities: 5—hula hooping; 4—shooting with slingshot; 3—paper airplanes; 2—paper basketballs; 1—doodling; 0—does nothing. Pretest scaling on 15 college students ordered these activities with respect to the degree of euphoria they represented. Arbitrary weights were assigned so that the wilder the activity, the heavier the weight. These weights are multiplied by an estimate of the amount of time the subject spent in each activity and the summed products make up the activity index for each subject. This index may be considered a measure of behavioral euphoria. It should be noted that the same between-condition relationships hold for the two components of this index as for the index itself.

The column labeled "Mean number of acts initiated" presents the data on the extent to which the subject deviates from the stooge's routine and initiates euphoric activities of his own.

On both behavioral indices, we find precisely the same pattern of relationships as those obtained with self-reports. Epi Mis subjects behave somewhat more euphorically than do Epi Ign subjects who in turn behave more euphorically than do Epi Inf subjects. On all measures, then, there is consistent evidence that a subject will take over the stooge's euphoric mood to the extent that he has no other explanation of his bodily state.

Again it should be noted that on these behavioral indices, Epi Ign and Epi Mis subjects are somewhat more euphoric than placebo subjects but not significantly so.

Anger: Self-report. Before presenting data for the anger conditions, one point must be made about the anger manipulation. In the situation devised, anger, if manifested, is most likely to be directed at the experimenter and his annoyingly personal questionnaire. As we subsequently discovered, this was rather unfortunate, for the subjects, who had volunteered for the experiment for extra points on their final exam, simply refused to endanger these points by publicly blowing up, admitting their irritation to the experimenter's face or spoiling the questionnaire. Though as the reader will see, the subjects were quite willing to manifest anger when they were alone with the stooge, they hesitated to do so on material (self-ratings of mood and questionnaire) that the experimenter might see and only after the purposes of the experiment had been revealed were many of these subjects willing to admit to the experimenter that they had been irked or irritated.

This experimentally unfortunate situation pretty much forces us to rely on the behavioral indices derived from observation of the subject's presumably private interaction with the stooge. We do, however, present data on the self-report scales in Table 4. These figures are derived in the same way as the figures presented in Table 2 for the euphoria conditions, that is, the value checked on the irritation scale is subtracted from the value checked on the happiness scale. Though, for the reasons stated above, the absolute magnitude of these figures (all positive) is relatively meaningless, we can, of course, compare condition means within the set of anger conditions. With the happiness-irritation index employed, we should, of course, anticipate precisely the reverse results from those ob-

TABLE 4

SELF-REPORT OF EMOTIONAL STATE IN THE ANGER CONDITIONS

Condition	N	Self-Report scales	Comparison	p
Epi Inf	22	1.91	Epi Inf vs. Epi Ign	.08
Epi Ign	23	1.39	Placebo vs. Epi Ign or Inf	*ns*
Placebo	23	1.63		

tained in the euphoria conditions; that is, the Epi Inf subjects in the anger conditions should again be less susceptible to the stooge's mood and should, therefore, describe themselves as in a somewhat happier frame of mind than subjects in the Epi Ign condition. This is the case; the Epi Inf subjects average 1.91 on the self-report scales while the Epi Ign subjects average 1.39.

Evaluating the effects of the injections, we note again that, as anticipated, Epi Ign subjects are somewhat less happy than Placebo subjects but, once more, this is not a significant difference.

Behavior. The subject's responses to the stooge, during the period when both were filling out their questionnaires, were systematically coded to provide a behavioral index of anger. The coding scheme and the numerical values attached to each of the categories have been described in the methodology section. To arrive at an "Anger index" the numerical value assigned to a subject's responses to the stooge is summed together for the several units of stooge behavior. In the coding scheme used, a positive value to this index indicates that the subject agrees with the stooge's comment and is growing angry. A negative value indicates that the subject either disagrees with the stooge or ignores him.

The relevant data are presented in Table 5. For this analysis, the stooge's routine has been divided into two phases—the first two units of his behavior (the "long" questionnaire and "What did I have for breakfast?") are considered essentially neutral revealing nothing of the stooge's mood; all of the following units are considered "angry" units for they begin with an irritated remark about the "bells" question and end with the stooge's fury as he rips up his questionnaire and stomps out of the room. For the neutral units, agreement or disagreement with the stooge's remarks is, of course, meaningless as an index of mood and we should anticipate no difference between conditions. As can be seen in Table 5, this is the case.

For the angry units, we must, of course, anticipate that subjects in the Epi Ign condition will be angrier than subjects in the Epi Inf condition. This is indeed the case. The Anger index for the Epi Ign condition is positive and large, indicating that these subjects have become angry, while in the Epi Inf condition the Anger index is slightly negative in value indicating that these subjects have failed to catch the stooge's mood at all. It seems clear that providing the subject with an ap-

TABLE 5

BEHAVIORAL INDICATIONS OF EMOTIONAL STATE IN THE ANGER CONDITIONS

Condition	N	Neutral units	Anger units
Epi Inf	22	+0.07	−0.18
Epi Ign	23	+0.30	+2.28
Placebo	22[a]	−0.09	+0.79

Comparison for anger units	p
Epi Inf vs. Epi Ign	$<.01$
Epi Ign vs. Placebo	$<.05$
Placebo vs. Epi Inf	*ns*

[a] For one subject in this condition the sound system went dead and the observer could not, of course, code his reactions.

propriate explanation of his bodily state greatly reduces his tendency to interpret his state in terms of the cognitions provided by the stooge's angry behavior.

Finally, on this behavioral index, it can be seen that subjects in the Epi Ign condition are significantly angrier than subjects in the Placebo condition. Behaviorally, at least, the injection of epinephrine appears to have led subjects to an angrier state than comparable subjects who received placebo shots.

Conformation of Data to Theoretical Expectations

Now that the basic data of this study have been presented, let us examine closely the extent to which they conform to theoretical expectations. If our hypotheses are correct and if this experimental design provided a perfect test for these hypotheses, it should be anticipated that in the euphoria conditions the degree of experimentally produced euphoria should vary in the following fashion:

$$\text{Epi Mis} \geqq \text{Epi Ign} > \text{Epi Inf} = \text{Placebo}$$

And in the anger conditions, anger should conform to the following pattern :

$$\text{Epi Ign} > \text{Epi Inf} = \text{Placebo}$$

In both sets of conditions, it is the case that emotional level in the Epi Mis and Epi Ign conditions is considerably greater than that achieved in the corresponding Epi Inf conditions. The results for the Placebo condition, however, are ambiguous for consistently the Placebo subjects fall between the Epi Ign and the Epi Inf subjects. This is a particularly troubling pattern for it makes it impossible to evaluate unequivocally the effects of the state of physiological arousal and indeed raises serious questions about our entire theoretical structure. Though the emotional level is consistently greater in the Epi Mis and Epi Ign conditions than in the Placebo condition, this difference is significant at acceptable probability levels only in the anger conditions.

In order to explore the problem further, let us examine the experimental factors identified earlier, which might have acted to restrain the emotional level in the Epi Ign and Epi Mis conditions. As was pointed out earlier, the ideal test of our first two hypotheses requires an experimental setup in which the subject has flatly no way of evaluating his state of physiological arousal other than by means of the experimentally provided cognitions. Had it been possible to physiologically produce a state of sympathetic activation by means other than injection, one could have approached this experimental ideal more closely than in the present setup. As it stands, however, there is always a reasonable alternative cognition available to the aroused subject—he feels the way he does because of the injection. To the extent that the subject seizes on such an explanation of his bodily state, we should expect that he will be uninfluenced by the stooge. Evidence presented in Table 6 for the anger condition and in Table 7 for the euphoria conditions indicates that this is, indeed, the case.

As mentioned earlier, some of the Epi Ign and Epi Mis subjects in their answers to the open-end questions clearly attributed their physical state to the injection, e.g., "the shot gave me the shivers." In Tables 6 and 7 such subjects are labeled "Self-informed." In Table 6 it can be seen that the self-informed subjects are considerably less angry than are the remaining subjects; indeed, they are not angry at all. With these self-informed subjects eliminated the difference between the Epi Ign and the Placebo

TABLE 6

THE EFFECTS OF ATTRIBUTING BODILY STATE TO THE INJECTION ON ANGER IN THE ANGER EPI IGN CONDITION

Condition	*N*	Anger index	*p*
Self-informed subjects	3	−1.67	*ns*
Others	20	+2.88	*ns*
Self-informed vs. Others			.05

conditions is significant at the .01 level of significance.

Precisely the same pattern is evident in Table 7 for the euphoria conditions. In both the Epi Mis and the Epi Ign conditions, the self-informed subjects have considerably lower activity indices than do the remaining subjects. Eliminating self-informed subjects, comparison of both of these conditions with the Placebo condition yields a difference significant at the .03 level of significance. It should be noted, too, that the self-informed subjects have much the same score on the activity index as do the experimental Epi Inf subjects (Table 3).

It would appear, then, that the experimental procedure of injecting the subjects, by providing an alternative cognition, has, to some extent, obscured the effects of epinephrine. When account is taken of this artifact, the evidence is good that the state of physiological arousal is a necessary component of an emotional experience for when self-informed subjects are removed, epinephrine subjects give consistent indications of greater emotionality than do placebo subjects.

Let us examine next the fact that consistently the emotional level, both reported and behavioral, in Placebo conditions is greater than that in the Epi Inf conditions. Theoretically, of course, it should be expected that the two conditions will be equally low, for by assuming that emotional state is a joint function of a state of physiological arousal and of the appropriateness of a cognition we are, in effect, assuming a multiplicative function, so that if either component is at zero, emotional level is at zero. As noted earlier this expectation should hold if we can be sure that there is no sympathetic activation in the Placebo conditions. This assumption, of course, is completely unrealistic for the injection of placebo does not prevent sympathetic activation. The experimental situations were fairly dramatic and certainly some of the placebo subjects gave indications of physiological arousal. If our general line of reasoning is correct, it should be anticipated that the emotional level of subjects who give indications of sympathetic activity will be greater than that of subjects who do not. The relevant evidence is presented in Tables 8 and 9.

As an index of sympathetic activation we shall use the most direct and unequivocal measure available—change

TABLE 7

THE EFFECTS OF ATTRIBUTING BODILY STATE TO THE INJECTION ON EUPHORIA IN THE EUPHORIA EPI IGN AND EPI MIS CONDITIONS

Epi Ign

	N	Activity Index	*p*
Self-informed subjects	8	11.63	*ns*
Others	17	21.14	*ns*
Self-informed vs. Others			.05

Epi Mis

	N	Activity Index	*p*
Self-informed subjects	5	12.40	*ns*
Others	20	25.10	*ns*
Self-informed vs. Others			.10

TABLE 8

SYMPATHETIC ACTIVATION AND EUPHORIA IN THE EUPHORIA PLACEBO CONDITION

Subjects whose:	N	Activity index	p
Pulse decreased	14	10.67	*ns*
Pulse increased or remained same	12	23.17	*ns*
Pulse decrease vs. pulse increase or same			.02

in pulse rate. It can be seen in Table 1 that the predominant pattern in the Placebo condition is a decrease in pulse rate. We shall assume, therefore, that those subjects whose pulse increases or remains the same give indications of sympathetic activity while those subjects whose pulse decreases do not. In Table 8, for the euphoria condition, it is immediately clear that subjects who give indications of sympathetic activity are considerably more euphoric than are subjects who show no sympathetic activity. This relationship is, of course, confounded by the fact that euphoric subjects are considerably more active than noneuphoric subjects—a factor which independent of mood could elevate pulse rate. However, no such factor operates in the anger condition where angry subjects are neither more active nor talkative than calm subjects. It can be seen in Table 9 that Placebo subjects who show signs of sympathetic activation give indications of considerably more anger than do subjects who show no such signs. Conforming to expectations, sympathetic activation accompanies an increase in emotional level.

TABLE 9

SYMPATHETIC ACTIVATION AND ANGER IN ANGER PLACEBO CONDITION

Subjects whose:	N^a	Anger index	p
Pulse decreased	13	+0.15	*ns*
Pulse increased or remained same	8	+1.69	*ns*
Pusle decrease vs. pulse increase or same			.01

[a] N reduced by two cases owing to failure of sound system in one case and experimenter's failure to take pulse in another.

It should be noted, too, that the emotional levels of subjects showing no signs of sympathetic activity are quite comparable to the emotional level of subjects in the parallel Epi Inf conditions (see Tables 3 and 5). The similarity of these sets of scores and their uniformly low level of indicated emotionality would certainly make it appear that both factors are essential to an emotional state. When either the level of sympathetic arousal is low or a completely appropriate cognition is available, the level of emotionality is low.

DISCUSSION

Let us summarize the major findings of this experiment and examine the extent to which they support the propositions offered in the introduction of this paper. It has been suggested, first, that given a state of physiological arousal for which an individual has no explanation, he will label this state in terms of the cognitions available to him. This implies, of course, that by manipulating the cognitions of an individual in such a state we can manipulate his feelings in diverse directions. Experimental results support this proposition for following the injection of epinephrine, those subjects who had no explanation for the bodily state thus produced, gave behavioral and self-report indications that they had been readily manipulable into the disparate feeling states of euphoria and anger.

From this first proposition, it must follow that given a state of physiological arousal for which the individual has a completely satisfactory explanation, he will not label this state in terms of the alternative cognitions available. Experimental evidence strongly supports this expectation. In those conditions in which subjects were injected with epinephrine and told precisely what they would feel and why, they proved relatively immune to any effects of the manipulated cognitions. In the anger condition, such subjects did not report or show anger; in the euphoria condition, such subjects reported themselves as far less happy than subjects with an identical bodily state but no adequate knowledge of why they felt the way they did.

Finally, it has been suggested that given constant cognitive circumstances, an individual will react emotionally only to the extent that he experiences a state of physiological arousal. Without taking account of experimental artifacts, the evidence in support of this proposition is consistent but tentative. When the effects of "self-informing" tendencies in epinephrine subjects and of "self-arousing" tendencies in placebo subjects are partialed out, the evidence strongly supports the proposition.

The pattern of data, then, falls neatly in line with theoretical expectations. However, the fact that we were forced, to some extent, to rely on internal analyses in order to partial out the effects of experimental artifacts inevitably makes our conclusions somewhat tentative. In order to further test these propositions on the interaction of cognitive and physiological determinants of emotional state, a series of additional experiments, published elsewhere, was designed to rule out or overcome the operation of these artifacts. In the first of these, Schachter and Wheeler (1962) extended the range of manipulated sympathetic activation by employing three experimental groups—epinephrine, placebo, and a group injected with the sympatholytic agent, chlorpromazine. Laughter at a slapstick movie was the dependent variable and the evidence is good that amusement is a direct function of manipulated sympathetic activation.

In order to make the epinephrine-placebo comparison under conditions which would rule out the operation of any self-informing tendency, two experiments were conducted on rats. In one of these Singer (1961) demonstrated that under fear inducing conditions, manipulated by the simultaneous presentation of a loud bell, a buzzer, and a bright flashing light, rats injected with epinephrine were considerably more frightened than rats injected with a placebo. Epinephrine-injected rats defecated, urinated, and trembled more than did placebo-injected rats. In nonfear control conditions, there were no differences between epinephrine and placebo groups, neither group giving any indication of fear. In another study, Latané and Schachter (1962) demonstrated that rats injected with epinephrine were notably more capable of avoidance learning than were rats injected with a placebo. Using a modified Miller-Mowrer shuttlebox, these investigators found that during an experimental period involving 200 massed trials, 15 rats injected with epinephrine avoided shock an average of 101.2 trials while 15 placebo-injected rats averaged only 37.3 avoidances.

Taken together, this body of studies does give strong support to the propositions which generated these experimental tests. Given a state of sympathetic activation, for which no immediately appropriate explanation is available, human subjects can be readily manipulated into states of euphoria, anger, and amusement. Varying the intensity of sympathetic activation serves to vary the intensity of a variety of emotional states in both rats and human subjects.

Let us examine the implications of these findings and of this line of thought for problems in the general area of the physiology of the emotions. We have noted in the introduction that the numerous studies on physiological differentiators of emotional states have, viewed en masse, yielded quite inconclusive results. Most, though not all, of these studies have indicated no differences among the various emotional states. Since as human beings, rather than as scientists, we have no difficulty identifying, labeling, and distinguishing among our feelings, the results of these studies have long seemed rather puzzling and paradoxical. Perhaps because of this, there has been a persistent tendency to discount such results as due to ignorance or methodological inadequacy and to pay far more attention to the very few studies which demonstrate *some* sort of physiological differences among emotional states than to the very many studies which indicate no differences at all. It is conceivable, however, that these results should be taken at face value and that emotional states may, indeed, be generally characterized by a high level of sympathetic activation with few if any physiological distinguishers among the many emotional states. If this is correct, the findings of the present study may help to resolve the problem. Obviously this study does *not* rule out the possibility of physiological differences among the emotional states. It is the case, however, that given precisely the same state of epinephrine-induced sympathetic activation, we have, by means of cognitive manipulations, been able to produce in our subjects the very disparate states of euphoria and anger. It may indeed be the case that cognitive factors are major determiners of the emotional labels we apply to a common state of sympathetic arousal.

Let us ask next whether our results are specific to the state of sympathetic activation or if they are generalizable to other states of physiological arousal. It is clear that from our experiments proper, it is impossible to answer the question for our studies have been concerned largely with the effects of an epinephrine created state of sympathetic arousal. We would suggest, however, that our conclusions are generalizable to almost any pronounced internal state for which no appropriate explanation is available. This suggestion receives some support from the experiences of Nowlis and Nowlis (1956) in their program of research on the effects of drugs on mood. In their work the Nowlises typically administer a drug to groups of four subjects who are physically in one another's presence and free to interact. The Nowlises describe some of their results with these groups as follows:

> At first we used the same drug for all 4 men. In those sessions seconal, when compared with placebo, increased the checking of such words as expansive, forceful, courageous, daring, elated, and impulsive. In our first statistical analysis we were confronted with the stubborn fact that when the same drug is given to all 4 men in a group, the N that has to be entered into the analysis is 1, not 4. This increases the cost of an already expensive experiment by a considerable factor, but it cannot be denied that the effects of these drugs may be and often are quite contagious. Our first attempted solution was to run tests on groups in which each man had a different drug during the same session, such as 1 on seconal, 1 on benzedrine, 1 on dramamine, and 1 on placebo. What does seconal do? Cooped up with, say, the egotistical benzedrine partner, the withdrawn, indifferent dramimine partner, and the slightly bored lactose man, the seconal subject reports that he is distractible, dizzy, drifting, glum, defiant, languid, sluggish, discouraged, dull, gloomy, lazy, and slow! This is not the report of mood that we got when all 4 men were on seconal. It thus appears that the moods of the partners do definitely influence the effect of seconal (p. 350).

It is not completely clear from this description whether this "contagion" of mood is more marked in drug than in placebo groups, but should this be the case, these results would certainly support the suggestion that our findings are generalizable to internal states other than that produced by an injection of epinephrine.

Finally, let us consider the implications of our formulation and data for alternative conceptualizations of emotion. Perhaps the most popular current conception of emotion is in terms of "activation theory" in the sense employed by Lindsley (1951) and Woodworth and Schlosberg (1958). As we understand this theory, it suggests that emotional states should be considered as at one end of a continuum of activation which is defined in terms of degree of autonomic arousal and of electroencephalographic measures of activation. The results of the experiment described in this paper do, of course, suggest that such a formulation is not completely adequate. It is possible to have very high degrees of activation without a subject either appearing to be or describing himself as "emotional." Cognitive factors appear to be indispensable elements in any formulation of emotion.

SUMMARY

It is suggested that emotional states may be considered a function of a state of physiological arousal and of a cognition appropriate to this state of arousal. From this follows these propositions:

1. Given a state of physiological arousal for which an individual has no immediate explanation, he will label this state and describe his feelings in terms of the cognitions available to him. To the extent that cognitive factors are potent determiners of emotional states, it should be anticipated that precisely the same state of physiological arousal could be labeled "joy" or "fury" or "jealousy" or any of a great diversity of emotional labels depending on the cognitive aspects of the situation.

2. Given a state of physiological arousal for which an individual has a completely appropriate explanation, no evaluative needs will arise and the individual is unlikely to label his feelings in terms of the alternative cognitions available.

3. Given the same cognitive circumstances, the individual will react emotionally or describe his feelings as emotions only to the extent that he experiences a state of physiological arousal.

An experiment is described which, together with the results of other studies, supports these propositions.

REFERENCES

AX, A. F. Physiological differentiation of emotional states. *Psychosom. Med.*, 1953, **15**, 433–442.

CANNON, W. B. *Bodily changes in pain, hunger, fear and rage.* (2nd ed.) New York: Appleton, 1929.

CANTRIL, H., & HUNT, W. A. Emotional effects produced by the injection of adrenalin. *Amer. J. Psychol.*, 1932, **44**, 300–307.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 114–140.

HUNT, J. McV., COLE, M. W., & REIS, E. E. Situational cues distinguishing anger, fear, and sorrow. *Amer. J. Psychol.*, 1958, **71**, 136–151.

JAMES, W. *The principles of psychology.* New York: Holt, 1890.

LANDIS, C., & HUNT, W. A. Adrenalin and emotion. *Psychol. Rev.*, 1932, **39**, 467–485.

LATANÉ, B., & SCHACHTER, S. Adrenalin and avoidance learning. *J. comp. physiol. Psychol.*, 1962, **65**, 369–372.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

MARAÑON, G. Contribution à l'étude de l'action émotive de l'adrénaline. *Rev. Francaise Endocrinol.*, 1924, **2**, 301–325.

NOWLIS, V., & NOWLIS, H. H. The description and analysis of mood. *Ann. N. Y. Acad. Sci.*, 1956, **65**, 345–355.

RUCKMICK, C. A. *The psychology of feeling and emotion.* New York: McGraw-Hill, 1936.

SCHACHTER, J. Pain, fear, and anger in hypertensives and normotensives: A psychophysiologic study. *Psychosom. Med.*, 1957, **19**, 17–29.

SCHACHTER, S. *The psychology of affiliation.* Stanford, Calif.: Stanford Univer. Press, 1959.

SCHACHTER, S., & WHEELER, L. Epinephrine, chlorpromazine, and amusement. *J. abnorm. soc. Psychol.*, 1962, **65**, 121–128.

SINGER, J. E. The effects of epinephrine, chlorpromazine and dibenzyline upon the fright responses of rats under stress and non-stress conditions. Unpublished doctoral dissertation, University of Minnesota, 1961.

WOLF, S., & WOLFF, H. G. *Human gastric function.* New York: Oxford Univer. Press, 1947.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1958.

WRIGHTSMAN, L. S. Effects of waiting with others on changes in level of felt anxiety. *J. abnorm. soc. Psychol.*, 1960, **61**, 216–222.

(Received February 17, 1961)

Psychological Review
1962, Vol. 69, No. 5, 400–414

ON THE QUANTIFICATION OF MILLER'S CONFLICT THEORY[1]

NORMAN H. ANDERSON

University of California, Los Angeles

This paper elucidates a mathematical model of potential usefulness in N. E. Miller's theory of conflict. Miller has discussed his theory most extensively in his provocative 1959 article and the present model is based fairly directly on that discussion. For simplicity of exposition, the model will be discussed in a rather restricted context. Only two particular experimental designs for a specific experimental task involving approach-avoidance conflict will be considered. However, the basic ideas and the techniques used have a considerably greater range of applicability.

In the experimental task to be considered, the human subject taps a key once upon presentation of a signal which is set to repeat at timed intervals. A goal, which may be represented by illuminating one of a row of lamps, moves one step closer to the subject at each tap of the key. When the last tap is made, the goal moves the last step, and the subject receives reward and punishment (e.g., coin and shock), either simultaneously or, in certain cases, on different trials. The dependent variable is response speed (reciprocal latency) as a function of distance from the goal. Each trial thus yields essentially a complete curve as the datum.

Provision for an overt retreat response could also be made and would possibly be useful in studying choice point vacillation. However, it will be assumed that only an approach or nonapproach response is permitted and, moreover, that the reward and punishment conditions are such that the subject reaches the goal on each trial. This latter restriction simplifies the presentation by avoiding, in particular, the more detailed statistical discussion needed when some response curves are incomplete.

The task described breaks up the goal response into a sequence of subresponses which has a number of advantages. The successive points on the response curve are more nearly physiologically independent than if a free continuous response were used. Since the goal moves, rather than the subject, the motor aspect of the successive subresponses should remain the same. In addition, various reinforcement schedules may be programed over the last several steps to the goal. Finally, the use of discrete subresponses provides simultaneous information on response probability and response strength which should be useful theoretically (Estes, 1950; Mueller, 1950; Spence, 1954).

MODEL

Assumptions

The model rests on the following assumptions:

A_1. The approach and avoidance gradients exist and are independent.

A_2. Response speed is proportional to the difference between the approach and avoidance gradients. Since the constant of proportionality represents only the choice of units on the measurement scales, it may conveniently be taken equal to unity. The assumption may then be written:

---

[1] A preliminary version of this paper was presented at the first annual meeting of the Psychonomic Society, September 1960, in Chicago.

$$S(x) = F(x) - G(x),$$

where x is distance from the goal, $F(x)$ is the approach gradient, $G(x)$ is the avoidance gradient, and $S(x)$ is response speed.

A_3. The approach and avoidance gradients can be written in the following form:

$$F(x) = af(x) + b + c,$$
$$G(x) = dg(x) + e + f,$$

where a, b are parameters depending on reward; d, e are parameters depending on punishment; c, f are constants in the given experimental situation; $f(x)$, $g(x)$ are completely arbitrary functions. The functions, $f(x)$ and $g(x)$, will be called the shape functions of the approach and avoidance gradients, respectively. The six parameters appearing in A_3 are assumed to vary with individuals and with trials. However it will be assumed here that no such variation occurs in the shape functions. This latter assumption will be discussed further under Application 2.

Of the three assumptions of the model, A_1 is the most stringent. The statement that the approach and avoidance tendencies exist is the main assertion about the psychological processes underlying the behavior. The statement that the gradients are independent is a brief way of saying that reward affects only the approach gradient and that punishment affects only the avoidance gradient. Of these two independence propositions, the first seems not unreasonable but the second should perhaps be viewed with greater reserve. Nevertheless, Miller does make these assumptions, both explicitly in Figures 7 and 9 of the cited reference (Miller, 1959), and implicitly in his procedure of comparing separately obtained approach and avoidance response curves. These assumptions will accordingly be used here. A somewhat weaker independence requirement will be discussed below.

The proportionality assumption, A_2, is more stringent than that made by Miller (1959) who requires only that the response measure be some monotone function of the difference between the gradients. It is, however, not entirely out of line with Hull-Spence theory (e.g., Logan, 1952; Spence, 1956) if the gradients are interpreted as reaction potentials. In any case, it will be seen below that only monotonicity is required if an appropriate experimental design is used.

Assumption A_3 is fairly weak since no restriction is placed on $f(x)$ and $g(x)$; these functions are entirely arbitrary and need not even be estimated for the proposed tests of the model. The given representation of the approach gradient assumes that the effect of reward is partly multiplicative and partly additive, as represented by the parameters a and b, respectively, and an analogous statement holds for punishment. The use of six parameters may seem peculiarly generous but the first test of the model will be completely parameter free, and the second test requires evaluation of only two parameters.

Before proceeding to the derivations, a few remarks on averaging are required since biases may result from differences between individuals or trials (Anderson, 1959, 1960, p. 85). For the present case, consider a given set of subjects and trials. Because of individual differences, subjects treated alike will in general have different gradients. Moreover, gradients will vary with trials because of learning, random fluctuation, or adaptation effects of various kinds. To the given set of subjects and trials therefore, corresponds a set or distribution of gradients of either type. Assumption A_1 implies that experimental conditions with say, the same punishment but different rewards will have the same distribution of avoidance gradients. This conclusion follows from A_1 under the present restriction that the subjects reach the goal on each trial, thus receiving both reward and punishment on each trial. Given this restriction, the assumption that reward and punishment act independently implies that the experimental conditions provide equivalent punishment treatment, and so yield identical distributions of avoidance gradients. In the same way, it is seen that experimental conditions with the same reward but different punishment have the same mean approach gradient. These two results are used to justify the use of average data in the following applications of the model.

Application 1

This application is based on a 2 × 2 design with two values each of reward and punishment. Consider a given subject receiving Reward i and Punishment j, and let $F_i(x)$ be the approach gradient, $G_j(x)$ the avoidance gradient, and $S_{ij}(x)$ the speed function for this subject on a given trial. Then by A_2:

$$S_{ij}(x) = F_i(x) - G_j(x). \quad [1]$$

This result holds for a given subject on a given trial. If a set of subjects and trials is considered, there will be a corresponding set or distribution of gradients. As shown above, the mean approach gradient, averaged over this distribution, will be the same for the two experimental conditions which have the same reward. Similarly, the mean avoidance gradient will depend only on the punishment involved, not on the reward. Since the mean of the difference on the right side of Equation 1 is the difference of the means, it follows that Equation 1 holds on the average over the set of subjects and trials under consideration. The quantities appearing in Equation 1 will accordingly be interpreted as means. When the four theoretical mean speed functions of Equation 1 are placed in the 2 × 2 table of the design, the following array is obtained:

$$\begin{array}{ll} F_1(x) - G_1(x) & F_1(x) - G_2(x) \quad [2] \\ F_2(x) - G_1(x) & F_2(x) - G_2(x). \end{array}$$

Inspection of this array shows that the difference between the two entries in the first row is equal to the difference between the two entries in the second row. This means, in analysis of variance terms, that the interaction of the row (reward) and column (punishment) variables is zero at each value of x. The model thus implies that the Reward × Punishment, and the Reward × Punishment × Distance interactions are zero in principle, and hence nonsignificant in practice. Consequently, the ordinary analysis of variance may be applied directly to the obtained speed curves in order to test the model.

This application of the model requires only A_1 and A_2 and is completely parameter free. Analogous results hold, of course, for any factorial design with reward and punishment as two of the factors.

Application 2

Consider now a situation with fixed punishment and g conditions of reward. The speed curve on a given trial for a given subject under Reward

i may be written:

$$S_i(x) = a_i f(x) + b_i + c - G(x) \qquad i = 1, 2, \cdots, g. \quad [3]$$

If Equation 3 is to be useful, it must be averaged over the distribution of gradients corresponding to a given set of subjects and trials. Now the average of the sum of terms on the right side of the equation is the sum of the average terms. As noted above, the mean $G(x)$ is the same in all reward conditions, and so also is the mean value of c. Since $f(x)$ was taken as fixed, the mean of the product, $a_i f(x)$, is just $f(x)$ times the mean a_i. Hence Equation 3 is valid if each term is interpreted as a mean and this will be done.

A rough test of the model for this application may be obtained by using Equation 3 to get expressions for, say, $S_2(x) - S_1(x)$, and $S_3(x) - S_1(x)$. These expressions may fairly easily be combined to show that the one difference is a simple linear function of the other difference:

$$S_3(x) - S_1(x) = A[S_2(x) - S_1(x)] + B, \quad [4]$$

where A and B are constants whose exact expressions in terms of the a_i and b_i are not of importance. The values of A and B could be estimated by using selected values of x, or by least squares. An assessment of goodness of fit could then be made by eye and this procedure, crude though it is, might be useful in many cases.

A rigorous statistical test for Application 2 may be obtained using the multivariate analysis of variance in the form of Fisher's (1938) test of coplanarity. For suppose Equation 3 is correct, and that the speed gradients are measured at, say, p values of x. Then it can readily be shown that the $p \times p$ between conditions sums of products matrix has Rank 2 ($g \geq 3$). The null hypothesis to be tested is then, in Rao's (1952) notation, that the smallest $p - 2$ latent roots of $B - \lambda W$ are zero, where B and W are the between and within conditions sum of products matrices, respectively. Rao gives numerical examples (pp. 364–375). The tests are approximate but evidently quite adequate in respect of significance level. The results of Marriott (1956) might be expected to yield somewhat more powerful tests. Unfortunately both procedures are somewhat tedious with a desk calculator.

Application 2 requires only half of A_1, namely, that reward does not affect the avoidance gradient. Also only half of A_3 is used since no restriction is placed on $G(x)$. The treatment of the case of fixed reward and variable punishment can be done in essentially the same way.

Since Application 2 uses the approximate approach gradient representation of A_3 the results are not exact, in contrast to those of the first application. It is interesting to note that the same results are obtained if a_i is assumed fixed within each reward condition and $f(x)$ is allowed to vary over trials and subjects. It would still be required that the mean $f(x)$ be the same in each condition, but the individual subjects could have sharply different approach gradient shape functions.

If both a_i and $f(x)$ were to be considered as variable over subjects and trials, each individual approach gradient would be represented exactly. However, if both a_i and $f(x)$ are variable, the mean of their product is the product of their means, plus the covariance of the two factors. In this case, Equation 3, interpreted in terms of means, would be biased because this covariance term is not included. Obversely, the covariance term indexes the adequacy of the approximation of the gradient representation of A_3. The bias in Equation 3 tends to cancel in the tests of goodness of fit and is not thought to be serious. However, it should be noted that the bias may be reduced by reducing the variability between trials and between subjects. Trial variability may be reduced by restricting attention to asymptotic data. The subject variability may be reduced by considering homogeneous blocks of subjects including, in the limit, single subjects.

Comments

If the model fails the experimental tests of the sort illustrated in the above two applications, one or more of the assumptions must be at fault.

The proportionality assumption, A_2, although presumably a reasonable first approximation, may well not be sufficiently exact for the quantitative tests under consideration. However, if the response measure is not proportional to the difference between the gradients, it should in any case be some strictly monotone function of this difference. From this it follows, by taking inverse functions, that there is some monotone function of the response measure which *is* proportional to the difference between the two gradients. Thus, if A_2 alone is at fault, it may in principle be rectified by employing this latter monotone function of the response measure in place of the original response measure. The only problem is whether this scaling of the response measure may be accomplished in practice. For the cited applications, the scaling presents no particular difficulty and the details are given in a subsequent section.

It thus follows that, if no scaling of the data satisfies the model, one of the other assumptions must be incorrect. In this case, since A_3 is quite weak and is used only in Application 2, A_1 would become suspect.

Assumption A_1 might be incorrect for two different reasons. First, the approach and avoidance gradients might give an inadequate representation of the processes underlying the behavior. Indeed, the present development, in contrast to that of Bower (1959) and Spence (1960, Ch. 22), does not give an explicit account of choice point vacillation. Further investigation of this aspect of the behavior would probably be quite useful as Miller (1959) has observed, and would be expected to guide any necessary revisions of the two gradient representation.

The second reason for rejection of A_1 would be nonindependence of the approach and avoidance gradients. If in fact reward and punishment affect both gradients, it will probably be necessary to include the learning processes explicitly in the model. This problem has not been investigated in any detail since it seems preferable to base the initial tests of the model on the simplest assumptions.

If the model is supported by the experimental tests, one's confidence in the assumptions would be increased, but it would still be useful to ask how much had been learned. Success in prediction gives only an indirect confirmation of the premises which yielded the prediction. Thus successful prediction would not prove A_1 because alternative assumptions might imply the same results as, indeed, the following example shows.

Consider Application 2 with reward and punishment given on separate trials. Suppose that the avoidance gradient has the same shape function as the approach gradient. Equation

3 may then be written:

$$S_i(x) = (a_i - d_i)f(x) + (b_i - e_i) + (c - f).$$

Suppose also that there is a decrement in the avoidance gradient on rewarded trials which results simply from the absence of punishment and so is independent of reward magnitude. This decrement would be represented by decreases in d_i and e_i in the equation above. Inspection of the equation shows that it is impossible to distinguish decreases in d_i and e_i from corresponding increases in a_i and b_i. It would thus be impossible to distinguish a simultaneous increase in $F(x)$ and decrease in $G(x)$ from a larger increase in $F(x)$ and no change in $G(x)$. Consequently, if it is also assumed that any effect of punishment on the reward gradient is independent of amount of reward, then the results derived in Application 2 hold also in the present case.

The example of the preceding paragraph, which seems to be a fairly reasonable possibility, indicates that the model will hold under a wider range of assumptions than those actually used in the derivation. This extra generality is a mixed blessing since, except for certain practical situations, prediction of behavior is only a means to the end of understanding behavior. However, the model does discriminate between variants of A_1 which are of interest. Moreover, it does not seem likely that the model predictions would be verified unless the measuring scale for the dependent variable were correct.

The estimation and interpretation of the shape functions and of the reward and punishment parameters also require discussion. The representation of the gradients used in A_3 is quite flexible and should give a good fit to any particular set of gradients. However, it is a representation of convenience, one which dissects the gradients in a mathematically powerful but possibly unpsychological way. Consequently, the reward and punishment parameters do not necessarily have any direct psychological interpretation. Whether or not they do cannot be determined without further empirical or theoretical analysis. Since the tests of the model do not require the estimation of these parameters, this problem is not considered here. Some comments on the interpretation of the parameters are given in a subsequent section.

Matters are somewhat clearer as regards the shape functions. If the model survives the test of Application 2, then Equation 3 may be used to estimate the shape function of the approach gradient. This shape function is only unique up to a multiplicative factor and an additive term so that it might perhaps be desirable to reduce to a standard form by setting $f(0) = 1$, $f(\infty) = 0$. How generally useful the resultant estimate would be is an open question. As Miller (1959) indicates, the gradients will be strongly influenced by the stimulus properties of the situation. There would thus be no reason to expect more than a rough similarity between shape functions from different experiments unless the procedures were carefully standardized. Within a given situation, however, estimation of the shape functions would be expected to be relevant to Miller's hypothesis that the avoidance gradient has a steeper slope than the approach gradient.

Another problem deserving comment arises because the present applications of the model are based on comparisons of different experimental conditions in which either the approach gradient or the avoidance gradient is the same in two or more conditions. If the present restriction that the subjects reach the goal on every trial is not met, the required comparability may not obtain. Thus subjects receiving a greater punishment might reach the goal less often than subjects receiving a lesser punishment. If so, the number of rewards would not be the same in the two conditions, the approach gradients would not be the same, and the above applications need not hold. In situations of this sort, there may be some justification for applying the model to the asymptotic data. For, by A_1, the same asymptotic reward gradients should be obtained, the lower rate of reward in the condition with greater punishment affecting only the rate at which the asymptote is reached.

In this connection, it is worth mentioning a potentially useful experimental procedure. The subject is run under a given reward and punishment until the behavior has stabilized and then told that on the next trial the reward will be increased. By A_1, this announcement should affect only the approach gradient. The avoidance gradients would be the same on the trial preceding as on the trial following the reward so that the data from these two trials would be comparable in the desired way.

Finally it should be noted that the model may be more generally applicable than has been indicated. It might be possible, depending on the experimental setup, to consider the two gradients as approach gradients with respect to two incompatible goals. In particular, some of the results of Application 1 hold even when there is but one step from start to goal. These results would be potentially applicable to situations involving two competing response tendencies.

LEARNING

Explicit incorporation of the learning processes into the model appears to present some difficulties. This section outlines one possible approach which should help set out the problems involved. The discussion is not essential to the rest of the paper and only a very brief sketch will be given.

For illustrative purposes, it is assumed that reward is given on a random proportion, π, of the trials, punishment being given on the remaining $1 - \pi$ of the trials. In the learning model to be considered, it is assumed that each gradient changes on each trial by a constant fraction of the possible change. This proportional change assumption is fairly common (Anderson, 1961; Bush & Mosteller, 1955; Estes, 1950; Hull, 1943) but it is here being applied to the whole gradient. Evidently it is necessary to allow the possibility of different constant fractions for the two gradients and for the two types of trials.

If reward is given on Trial n, the gradients on Trial $n + 1$ are:

$$\begin{aligned} F_{n+1}(x) &= F_n(x) \\ &\quad + \theta_1[F_U(x) - F_n(x)], \quad [5] \\ G_{n+1}(x) &= G_n(x) \\ &\quad + \theta_2[G_L(x) - G_n(x)]. \end{aligned}$$

If punishment is given on Trial n, the gradients on Trial $n + 1$ are:

$$\begin{aligned} F_{n+1}(x) &= F_n(x) \\ &\quad + \theta_1'[F_L(x) - F_n(x)], \quad [6] \\ G_{n+1}(x) &= G_n(x) \\ &\quad + \theta_2'[G_U(x) - G_n(x)]. \end{aligned}$$

Here θ_1, θ_2, θ'_1, θ'_2 are the four learning rates; subscripts n and $n + 1$ index the gradients on Trials n and $n + 1$; subscripts U and L index the upper and lower limit gradients, quantities analogous to the limit points of the linear operators of Bush and Mosteller (1955). The quantities in brackets thus denote the possible change on Trial n, and the equations state simply that the actual change is a constant fraction of the possible change.

Although the learning model is conceptually simple, its complete analysis would involve the estimation of four learning rates and four limit gradients. It seems reasonable to suppose that two of the limit gradients, $F_L(x)$ and $G_L(x)$, are identically zero, and this will be done in what follows. Further restrictions on the parameters seem problematical so that the estimation problem would be difficult. However, analysis of the sequential dependencies, that is, of the trial-to-trial increments and decrements in the response measure, does yield some information which may be useful.

On the basis of the above formulation, the speed increment on rewarded trials, denoted by $C_R(x)$, can be shown to be:

$$C_R(x) = \theta_1[F_U(x) - F_n(x)] + \theta_2 G_n(x).$$

On punished trials, the speed decrement is:

$$C_P(x) = \theta'_1 F_n(x) + \theta'_2[G_U(x) - G_n(x)].$$

The sum of the increment and decrement is thus:

$$C_R(x) + C_P(x) = \theta_1 F_U(x) + \theta'_2 G_U(x) + (\theta'_1 - \theta_1) F_n(x) - (\theta'_2 - \theta_2) G_n(x). \quad [7]$$

The left side of the equation may be obtained from the data. For a given set of trials, $C_R(x)$ would be estimated by the speed increment averaged over all rewarded trials; similarly $C_P(x)$ would be estimated by the speed decrement averaged over all punished trials. $F_n(x)$ and $G_n(x)$ would correspondingly be replaced by their theoretical averages over the given set of trials.

Now $F_n(x)$ and $G_n(x)$ depend on the value of π; however, the other quantities on the right side of Equation 7 are independent of π. If reward and punishment had equivalent effects so that $\theta'_1 = \theta_1$, and $\theta'_2 = \theta_2$, then the last two terms on the right side vanish. In this case, the model implies that $C_R(x) + C_P(x)$ has the same values for different groups run under different values of π. In this case, therefore, the sequential dependency provides an exact test of the model. Since the last two terms of Equation 7 involve learning rate differences, the sequential dependency should be approximately independent of π as long as the effects of reward and punishment are not too disparate.

If punishment is fairly strong compared to reward, it would be expected that both $\theta'_1 > \theta_1$ and $\theta'_2 > \theta_2$. Then, since $F_n(x)$ is an increasing function of π, and $G_n(x)$ is a decreasing function of π, the sequential dependency should show modest increases as π is increased. Conversely, if punishment is fairly weak compared to reward, it would be expected that $\theta'_1 < \theta_1$, and $\theta'_2 < \theta_2$. The sequential dependency would then show modest decreases as π is increased.

Related results may be obtained from the higher order sequential dependencies, and one of these deserves particular mention. Let $C_{PR}(x)$ denote the change in response caused by a pair of successive trials, the first of which is punished and the second rewarded. Similarly, let $C_{RP}(x)$ be the speed change caused by a rewarded and a punished trial given in that order. The differences between these quantities is always positive and indeed:

$$C_{PR}(x) - C_{RP}(x) = \theta_1\theta'_1 F_U(x) + \theta_2\theta'_2 G_U(x). \quad [8]$$

The difference on the left side of this equation is to be estimated from the data in the same way as mentioned above. The equation asserts that this difference is independent of π. This result affords a rigorous nonparametric test of goodness of fit. It should also be noted that the difference of Equation 8 varies directly with the learning rates thus yielding a qualitative check of the model.

The above results are less complete than would be desired but they do permit an assessment of the model. Although in principle the sequential dependencies provide the information for a complete solution to the parameter estimation problem (Anderson, 1959), practical difficulties are great. Given two types of gradient and two types of reinforcing stimuli, however, there are necessarily a considerable number of quantities to be estimated. Consequently, it may be useful to consider the possibility of a representation of the conflict situation which is mathematically more tractable even if possibly less realistic. It should be noted, however, that the trial-to-trial increments and decrements in response strength are the basic quantities in the study of learning, and their investigation should prove valuable under any circumstances.

SCALING THE RESPONSE MEASURE

This section shows how the scaling of the response measure mentioned above may be accomplished. The method is general in that it applies to any continuous response measure. However, the scaling equations are fairly specific to the particular experimental design employed so that a general presentation is not possible. Consequently, only a brief sketch indicating how the method would be used in the above two applications will be given. Statistical details, such as the loss of *df* in estimating the auxiliary scaling parameters, will not be considered.

The basic idea of the present method is to use the theoretical equations directly in the scaling procedure. This approach may be more generally applicable. To the author, it seems considerably more attractive than Hull's (Hull, Felsinger, Gladstone, & Yamaguchi, 1947) quantification procedure because the psychophysical-psychometric assumptions on which the latter is based are not closely related to the form of the theoretical behavior laws proposed for investigation.

If scaling is possible in principle, then, as noted above, there is some monotone function, $\mathfrak{M}$, of the observed response measure, $S(x)$, which does satisfy the proportionality assumption, A_2. Expanding $\mathfrak{M}[S(x)]$ in a Maclaurin series yields:

$$\mathfrak{M}[S(x)] = S(x) + \alpha_2 S^2(x) + \alpha_3 S^3(x) + \cdots. \quad [9]$$

The constant term in the series expansion does not appear since it is required that $\mathfrak{M}(0) = 0$. Moreover, it is no restriction on present purposes to take the coefficient, α_1, of $S(x)$ to be unity and this has been done. The number of terms of the series which would be required depends on how well the original scale approximates the final scale; only the first three terms of the series will be used here since it would seem that that should be adequate for most work.

Consider first Application 1 and for convenience of notation let:

$$P_n(x) = S_{11}{}^n(x) - S_{12}{}^n(x) - S_{21}{}^n(x) + S_{22}{}^n(x), \qquad n = 1, 2, \cdots$$

The assumption that $S(x)$ is proportional to the difference between the gradients is equivalent to the requirement that the Reward × Punishment interactions of Equation 2 be zero. These interactions will be zero when the difference between the entries in the first row equals the difference between the entries in the second row for each value of x. In present notation, this condition means concisely that $P_1(x) = 0$ for each x.

If $\mathfrak{M}[S(x)]$, rather than $S(x)$ itself, is to be proportional to the difference between the gradients, a condition analogous to $P_1(x) = 0$ must hold. If $S(x)$ in Equation 2 is replaced by $\mathfrak{M}[S(x)]$ as represented in the series expansion of Equation 9, this condition is seen to be:

$$P_1(x) + \alpha_2 P_2(x) + \alpha_3 P_3(x) = 0.$$

Scaling is thus equivalent to choosing α_2 and α_3 so that this last equation is true for each x. In practice, of course, this would not be possible because of sampling error. Consequently, α_2 and α_3 are chosen so as to minimize the interaction, using the method of least squares (e.g., Lewis, 1960). The sum of squared deviations is $\Sigma_x[P_1(x) + \alpha_2 P_2(x) + \alpha_3 P_3(x)]^2$. This sum is to be differentiated with respect to α_2 and α_3 and these derivatives set equal to zero. The following two linear equations result:

$$\Sigma_x P_1(x)P_2(x) + \alpha_2 \Sigma_x P_2^2(x) + \alpha_3 \Sigma_x P_2(x)P_3(x) = 0,$$

$$\Sigma_x P_1(x)P_3(x) + \alpha_2 \Sigma_x P_2(x)P_3(x) + \alpha_3 \Sigma_x P_3^2(x) = 0.$$

Since the values of the $P_n(x)$ are obtainable from the raw data, these equations are easily solved for α_2 and α_3. With these values of α_2 and α_3, the observed responses would be scaled using Equation 9 and the analysis of variance applied directly to the scaled data. If the interaction remains significant in the scaled data, then the scaled response measure does not satisfy the proportionality assumption. If additional terms of the series expansion of Equation 9 were used, say $N + 1$ altogether, the scaling method would yield the following N linear equations in N unknowns:

$$\Sigma_x \Sigma_s \alpha_s P_s(x) P_t(x) = 0, \qquad t = 2, 3, \cdots, N + 1.$$

The scaling technique for Application 2 follows similar lines. Let:

$$R_n(x) = S_3{}^n(x) - AS_2{}^n(x) + (A - 1)S_1{}^n(x), \quad n = 1, 2, \cdots.$$

From Equation 4, the requirement of the model in terms of the original response measure is that $R_1(x) - B = 0$. In terms of the scaled measure, this requirement becomes:

$$R_1(x) + \alpha_2 R_2(x) + \alpha_3 R_3(x) - B = 0.$$

It is thus required to minimize $\Sigma_x[R_1(x) + \alpha_2 R_2(x) + \alpha_3 R_3(x) - B]^2$. This minimization may be accomplished by using least squares to obtain four equations in four unknowns, α_2, α_3, A, B. The equations, which are not given here, are unfortunately mildly nonlinear in the unknowns so that a process of successive approximation would be required for their solution.

Scales obtained from the above procedure are subject to a bias since the mean of the transformed scores will not in general be exactly equal to the transform of the mean score. The amount of bias would be expected to be negligible unless $\mathfrak{M}$ is considerably nonlinear. If $\mathfrak{M}$ is badly nonlinear, it would probably be advisable to iterate the scaling procedure in order to converge on an unbiased scale.

If the $\mathfrak{M}$ obtained from the scaling procedure is nonmonotone, then one must of course reject the possibility of a theoretically adequate scaling, and hence also the basic theory. However, it is apparent that the scaling technique may be rather too obliging in yielding too readily, in a given experimental situation, a transformation of the observed response measure which satisfies the model. Nevertheless, there is a requirement of generality implicit in the notion of a measurement scale; it must be applicable to a somewhat wider class of situations than that on which the scaling is based. Essentially this problem has been discussed by Seward (1955) as a requirement of constancy of the intervening variable. It suffices here to note that this requirement of generality also means that the scaling may be done in a convenient situation. The resultant scale may then be used to test the model in situations in which the solution of the scaling equations would be difficult or impossible.

In this connection, it should be observed that the scaling procedure itself may be employed in such a way as to give a consistency check. If the distance variable is divided into subranges, the scaling procedure may be applied to the data of each separately. Theoretically, the same scale should be obtained in each subrange so that comparison of the separately obtained scales provides a test of the theory.

As has been seen, the logic of the present scaling technique consists in using the postulated behavior laws to induce a scaling on the dependent variable. In principle, this same approach could be applied to any theory which postulates quantitative laws of behavior although the detailed procedure would depend on the form of those laws and on the experimental design employed. The resulting scales would of course be correct only within the context of the postulated laws. However, they would have the considerable merit of being directly relevant to those postulates and of being directly associated with the development of the theory.

Finally, it should be noted that possession of a theoretically adequate scale of the dependent variable opens up the possibility of scaling the underlying variables. Miller's (1959) suggestion that a change in drive raises or lowers the whole gradient may be taken as one example. The difference between the scaled response before and after the change in drive should then represent the effect of that change in a theoretically adequate way. This would, for instance, give a basis for equating drives induced by different experimental operations.

POSSIBLE RELATIONS TO HULL-SPENCE THEORY

The purpose of this section is to explore one possible way in which the present formal model might be interpreted within a more general quantitative behavior theory. Such an interpretation is not necessary, of course, since the model is self-contained, but there are at least two reasons for so proceeding. An articulation of the model with an existing theory would presumably help to define the meaning of the model parameters. More importantly, such articulation would be expected to shed light on the validity of the assumptions and to indicate possible directions of change if change should prove necessary.

It is appropriate to restrict the discussion to general Hull-Spence theory. The treatment of this theory may seem somewhat casual but it should be realized that the intent is to indicate that an interpretation of the model within the theory may be possible and useful.

The following notation will be used. E, H, D, and K refer to reaction potential, habit strength, drive, and incentive motivation, respectively. Superscripts $+$ and $-$ index quantities associated with the approach and nonapproach responses, respectively. Thus the approach gradient is denoted by $E^+(x)$, and the avoidance gradient by $E^-(x)$.

Attention will be restricted to the following experimental situation. The subject is trained with a fixed reward and punishment until his behavior has stabilized. Theoretically, the values of $E^+(x)$, $E^-(x)$, $H^+(x)$, and $H^-(x)$ will then have reached asymptote. He is then given one of two instructions, either that the reward on the next trial will be greater (or less), or else that the punishment on the next trial will be greater (or less). Subscripts 1 and 2 will refer to quantities on the trial preceding and on the trial following this instruction, respectively.

Case a: Effect of Changed Reward

When the behavior has stabilized, the reaction potential difference corresponding to the difference between the gradients may, within Hull's theory, be written as:

$$E^+(x) - E^-(x) = K^+DH^+(x) - E^-(x). \quad [10]$$

The effect of the instruction will presumably be on K^+. Denote the new incentive motivation by αK^+ where α is constant. The reaction potential difference after the instruction is then:

$$E^+(x) - E^-(x) = \alpha K^+DH^+(x) - E^-(x). \quad [11]$$

To relate this to the model, set $f(x) = H^+(x), a_1 = K^+D, a_2 = \alpha K^+D$, $b = c = 0$. The Hullian equations are thus seen to be consistent with the model, at least for this situation. Conversely, this application of Hullian theory is seen to justify A_1 as regards the approach gradient.

An analogous treatment in terms of Spence's (1956, p. 151) theory is more problematical. Apparently, K^+ should be considered as a function of x, but H^+ should be taken as independent of x. Although it would seem that D should be considered as dependent on x, it will be assumed here that D is the same at all distances from the goal, an assumption which will be given some justification below. These assumptions yield, analogous to Equation 10:

$$E^+(x) - E^-(x) = [K^+(x) + D]H^+ - E^-(x). \quad [12]$$

The effect of the instruction is again presumably to change the incentive motivation and it is assumed that the new incentive motivation is related to the old by the expression, $\alpha K^+(x) + \beta$, where α and β are constants. It will be noted that this assumption includes a number of special cases: additive, with $\alpha = 1$; multiplicative, with $\beta = 0$; and exponential growth with $0 \leq \alpha \leq 1$, $\beta = (1 - \alpha)K_M{}^+$, where $K_M{}^+$ is the maximum value of the incentive motivation. This yields, analogous to Equation 11:

$$\begin{aligned} E^+(x) - E^-(x) &= [\alpha K^+(x) + \beta + D]H^+ \\ &\quad - E^-(x) = \alpha H^+K^+(x) \\ &\quad + \beta H^+ + DH^+ - E^-(x). \quad [13] \end{aligned}$$

Setting $f(x) = K^+(x)$, $a_1 = H^+$, $a_2 = \alpha H^+$, $b_1 = 0$, $b_2 = \beta H^+$, and $c = DH^+$, it is seen that this interpretation of Spence's (1956) theory gives a result which is consistent with the model, at least for the given experimental situation. In particular, it justifies A_1 as far as the approach gradient is concerned.

Case b: Effect of Changed Punishment

In this case, the problem is whether an analysis of E^- can be made which is similar to the analysis of E^+ given in Case a. This does not seem possible within either Hull's or Spence's theory. For a threat of increased punishment would presumably increase the drive, D, and change the drive stimulus, S_D. The increased D would, as is well-known, theoretically act to increase the reaction potential difference, and hence also the speed of response, at all distances up to that point at which the subject no longer approaches. The changed drive stimulus would presumably have the opposite effect, of course, but unless the changes in D and S_D exactly balanced out in their effect on the approach tendency, the assumption of independence of the approach and avoidance gradients would be incorrect. It would thus appear that Miller's (1959) theory is, in this respect, inconsistent with the theories of Hull and of Spence.

There is, however, a possible alternative approach. This is to postulate a negative incentive motivation, K^-, which acts on H^- in the same way that K^+ acts on H^+, and to assume that the effect of the punishment instruction is on K^-. The theory would then be symmetrical with respect to approach and avoidance so that the analysis of E^+ in Case a would carry over directly to the analysis of E^- and exactly analogous conclusions would follow.

The concept of negative incentive motivation as used here would seem to merit some consideration. Such a concept has been used by Logan (1960, p. 222), and it is consistent with Miller's (1959) stimulus interpretation of drive. Indeed, if K^- were introduced, the concept of generalized drive, D, might be no longer necessary.

It should also be noted that, even if H^+ and H^- were constant at asymptote, K^+ and K^- might well still show trial-to-trial variation as a function of the sequence of reinforcing stimuli. If so, the learning formulation given in an earlier section would be directly applicable to the asymptotic data.

DISCUSSION

Any psychological theory has four aspects. First, a symbolic representation of the organism or, in less nominal terms, a delineation of the psychological processes underlying the behavior. Second, a set of laws relating these processes to their causal environmental determinants. Third, a set of combination rules which state how the underlying processes combine or coact to produce the observable behavior. Fourth, a set of measurement scales for the dependent and independent variables, and for the underlying processes. This fourth aspect is, of course, implicit in the first three but it will be useful to consider it in its own right.

The principal concern of the present model is with the last two aspects. The basic combination rule of the model is the assumption, A_2, that the overt response is proportional to the difference between the strengths of the two underlying processes, namely, the approach and avoidance gradients. By applying this assumption, under the general guidance of Assumption A_1, to various experimental situations, precise theoretical predictions are derived and serve as a test of the model. An important by-product, one which often results from a quantitative theoretical analysis, is the illumination of various experimental and statistical considerations of the kind discussed above which are involved in the detailed application of the model. This by-product would seem the more valuable in that the relevance of a number of these considerations does not appear to be limited to the specific conflict model employed here.

It will be noted, however, that the exactness of prediction conferred by A_2 is not entirely satisfactory because of the arbitrary choice of a measuring scale for the dependent variable. Although response speed may be a reasonable approximation to a theoretically adequate scale, it would be optimistic to believe that it will be exact. Without a guarantee of A_2, however, it would be difficult to decide whether a deviation from prediction signals a basic flaw in the theoretical structure, or simply an inappropriate choice of scale of measurement. The difficulty is resolved by the scaling technique which yields a theoretically adequate scale of the dependent variable, i.e., a scale which is correct within the model employed.

The present development has little direct relevance to the first two aspects of theory listed above. The assumed underlying processes, namely, the approach and avoidance tendencies, are just those postulated by Miller (1959). Moreover, the scope of the model is narrow and formal in contrast to the substantive investigations of Miller and his colleagues which have been largely directed at the establishment of a set of causal laws, represented in particular by the assumption of gradient independence. Nevertheless, the model does bear indirectly on these two aspects. For, as has been seen, the success or failure of the model rests finally on Assumption A_1. If the model is rejected, then the assumption of gradient independence must be abandoned, or a different representation of the organism must be sought. In addition, it would seem that a theoretically adequate scale of the dependent variable is useful in defining the underlying processes and prerequisite to establishing scales for them.

Although much of the glamor of mathematical models seems to be associated with their predictive precision, this is not a good in itself. Prediction is usually only a means to the end of understanding behavior. Moreover, when there are several underlying processes, even a quantitative attack may yield only qualitative results. The advantage of a mathematical over a verbal formulation is that the former brings the basic ideas of the theory more quickly and more definitely to test. It is in this light that the present work should be viewed.

SUMMARY

This paper presents a mathematical model based on N. E. Miller's (1959) conflict theory and requiring minimal assumptions. An experimental situation is described and exact predictions are derived for it. A technique for obtaining a theoretically correct measurement scale for the dependent variable is also given. Finally, possible relations of the model to Hull-Spence theory are discussed.

REFERENCES

ANDERSON, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Pp. 248–264.

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

ANDERSON, N. H. Two learning models for responses measured on a continuous scale. *Psychometrika*, 1961, **26**, 390–403.

BOWER, G. H. Choice-point behavior. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Pp. 109–124.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

FISHER, R. A. The statistical utilization of multiple measurements. *Ann. Eugen., London*, 1938, **8**, 376–386.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L., FELSINGER, J. M., GLADSTONE, A. I., & YAMAGUCHI, H. G. A proposed quantification of habit strength. *Psychol. Rev.*, 1947, **54**, 237–254.

LEWIS, D. *Quantitative methods in psychology.* New York: McGraw-Hill, 1960.

LOGAN, F. A. Three estimates of differential excitatory tendency. *Psychol. Rev.*, 1952, **59**, 300–307.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

MARRIOTT, F. H. C. Tests of significance in canonical analysis. *Biometrika*, 1956, **39**, 58–64.

MILLER, N. E. Liberalization of basic S-R concepts. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

MUELLER, C. G. Theoretical relationships among some measures of conditioning. *Proc. Nat. Acad. Sci.*, 1950, **36**, 123–130.

RAO, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

SEWARD, J. P. The constancy of the *I-V*: A critique of intervening variables. *Psychol. Rev.*, 1955, **62**, 155–168.

SPENCE, K. W. The relationship of response latency and speed to the intervening variables and *N* in S-R theory. *Psychol. Rev.*, 1954, **61**, 209–216.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs: Prentice-Hall, 1960.

(Received January 9, 1961)

Psychological Review
1962, Vol. 69, No. 5, 415–427

FROM ASSOCIATION TO STRUCTURE[1]

GEORGE MANDLER

University of Toronto

The dialogue between stimulus-response theories and cognitive theories has antecedents going back at least as far as the beginnings of scientific psychology. Protestations that the conflict is imaginary or merely semantic are unlikely to eliminate fundamental theoretical differences. To argue that we are dealing with the same data regardless of theoretical predilections (e.g., Kendler, 1952) would sound strangely incongruous in the context of other sciences. The question "What is the center of the universe?" was not a theoretical blind alley, and neither is the question "What is learned?" as Campbell (1954) and Rozeboom (1958) have shown rather convincingly. In the present paper we shall examine briefly the major differences between the two positions with regard to the definition of response or output and then proceed to illustrate a possible locus of transition from associative to cognitive or structural processes.

The distinction between associationist and cognitive views may be described as a difference between a topographic and a categorical response definition. While the associations of the associationistic schools are usually considered to be between discrete events in the environment (the stimuli) and topographically well delimited and discrete events associated with the organism (the responses), the cognitive theorist is more likely to talk about formal categories of behavior, giving relatively less attention to the topographic aspects of the behavior classed within those categories. The important questions are: Do organisms learn generalizable, but discrete, responses in specific situations or are rules of behavior, maps, or schemata laid down which connect various behaviors and environmental inputs? Do organisms learn what to "do," or do they learn "what leads to what"? While the distance between the two views has been reduced somewhat by suggestions such as Logan's (1960), who aggregates responses as a function of differential reinforcement, the problem of response definition still appears to be basic.

The so-called cognitive theories belong to a much larger class of theories which have claimed to be "structural," including not only the gestalt school and its heirs but also the speculations of Piaget, Bartlett, and Hebb, among others. What these positions have in common is a postulation of the organization of behavior not derivable from any combination or association of stimulus-response links.

Rather than reserving the construct of "structure" for this kind of theory, we would prefer a more general statement of the structural position. *Structures are temporal and probabilistic linkages of inputs and behavior which are available in functional units.* These units—broadly conceived—may be as simple as a reflex arc, as associationistic as a habit-family hierarchy, or as cognitive as a means-end relationship. The output of a structure may be topographically limited or categorically defined; the question to be addressed concerns the conditions under which structures are developed and the relation between structures which mediate topographically limited behavior and those which mediate behavior described in categorical form.

---

[1] This paper was prepared in connection with Research Grant M-4852 from the National Institute of Mental Health, United States Public Health Service. The section on overlearning and transfer was presented at the American Psychological Association Symposium on Transfer of Training in September 1957. The initial version of the paper was prepared while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California. He is indebted to William Kessen, who contributed not only valuable critical goads but also additions of substance, and to Jean Mandler for her many suggestions.

The major notion to be presented in this paper and applied to a limited set of phenomena is the proposition that an associationist and a cognitive view of behavior may work side by side—that they are compatible. More specifically, it will be maintained here that there is now enough evidence available to suggest that, in some limited cases, cognitive characteristics of the organism may be developed out of associationist processes.

FROM ASSOCIATION TO COGNITION

The broad area of evidence to which we will appeal in support of such a proposition covers the warm-up effect and the development of learning sets. More specific evidence will be developed out of a series of studies, strewn through the literature, on the effects of overlearning on transfer of training.

We can briefly make some distinctions between three types of effects:

The warm-up effect. Prior experience in the particular setting of an experiment results in an increasing facilitation on the criterion task as a function of amount of such experience (Mandler, 1956; Thune, 1950). Thus Irion (1948) described the warm-up effect in general; while Thune (1950) demonstrated that both color-guessing and paired-associate learning were equally effective in facilitating learning of a test list. Hamilton (1950) suggested that these effects may be relatively transitory in nature.

Learning to learn. When a subject is given repeated and related tasks, there is an increasing facility in the performance of these tasks. While this effect has been most dramatically demonstrated by Harlow (1949), Bair (1902) in a classic study was able to conclude that "learning to do one thing gives us the capacity for learning to do other things" (p. 45). Most recently, Duncan (1960) showed that performance improved markedly as a function of number of prior tasks.

Overlearning and transfer. High degrees of learning of a particular response lead to positive transfer to an incompatible response, even though lower levels of learning produce negative transfer (cf. Reid, 1953).

These three general effects all have one characteristic in common, *practice on one task may have a facilitating effect on another task even though the behavior required in the two tasks is topographically different, and may even be incompatible.*

To what extent can current associationist stimulus-response theories accommodate to these phenomena? As far as the warm-up effect is concerned, it seems to be reasonable to assume that within broad limits a warm-up task permits the subject to "assume more rapidly the postural and 'attentive' adjustments necessary for optimal performance" (Thune, 1950, p. 438). However, when we deal with learning sets and overlearning, a topographically limited stimulus-response theory sometimes fails to do justice to the phenomena.

The general position to be adopted here in respect to learning sets and overlearning is similar to that espoused originally by Harlow (1949) and expanded subsequently (1959).[2] Harlow suggested that the formation of hypotheses and insight is a function of the continuous development of learning sets, rather than the discontinuous appearance and replacement of cognitive structures. The writer has previously suggested (Mandler, 1954a) that during overlearning, i.e., continuous errorless performance, there are developed symbolic analogues—we now prefer to call them analogic structures—of the overt responses. The process unfolds in the following manner: First, the organism makes a series of discrete responses, often interrupted by incorrect ones. However, once errors are dropped out and the sequence of behavior becomes relatively stable—as in running a maze, speaking a word, reproducing a visual pattern—the various components of the total behavior required in the situation are "integrated." Integration refers to the fact that previously discrete parts of a sequence come to behave functionally as a unit; the whole sequence is elicited as a unit and behaves as a single component response has in the past; any part of it elicits the whole sequence. This process of integration has been called "response learning" (Mandler, 1954a) and may be applied to any consistent responding of whatever topography and complexity. Thus, pressing a button is integrated as much as running a maze or speaking a sentence. Once a response sequence has been integrated and acts as a unit, it develops a structural representation, a "central" analogue of this new response unit which can function independently of the overt response sequence.

The notion of integration deals in part with the problem of defining units of behavior. When complex behavior is built up from previously disparate units, much trial and error behavior may characterize the slow development of the new, larger unit. The fact that the components may themselves have been overlearned—and structurally represented—lends a peculiarly "cognitive" flavor to the explicit or implicit hunt and peck search for the appropriate components. However, once the units have been selected and the new sequence has been integrated, it tends to be elicited as a whole and has a quasi-automatic appearance in the way it is evoked by the appropriate situational cues. It is this development from unsure steps to the well integrated sequence that characterizes learning to drive a car, to type a word, or to run a maze. Such a process is related to Miller's (1956) discussion of the chunking of behavior and probably led to Tolman's (1933) conclusion that "the Sign-gestalt formula is appropriate only for the early stages of learning and not for the later ones" (p. 254).

Analogic structures permit covert trial and error behavior, i.e., cognitive manipulation of previously established behavior. In this sense, the analogic representation of a prior behavior sequence is one possible "hypothesis" to be applied to a particular situation. Given many such structures which are relevant to a situational input, the several structures will occur seriatim and covertly until an appropriate one is expressed behaviorally. If it is incorrect, a new structure may occur, thus giving the appearance of an entirely discontinuous process, of shifting from hypothesis to hypothesis. We say "appearance" of discontinuity because a continuous process prior to the appearance of cognitive structures in fact gives rise to them. Structures are developed on the basis of associationist stimulus-response relationships but, once established, enable the organism to behave "cognitively." In Harlow's (1959) words: ". . . it is the amount of intraproblem learning which determines the interproblem transfer" (p. 502).

---

[2] The point of view presented here has familial ties with Hebb's (1949) theory, with some of Piaget's (1952) ideas, and with a recent essay by Miller, Galanter, and Pribram (1960).

In the learning set studies, the problem is not one of covert trial and error, i.e., the choice between two incompatible or antagonistic responses, but rather one of positive transfer from one problem to the other when the tasks have formal similarities, but no overlap in content. What will be suggested here—and this is one of the reasons why we should move from specific symbolic analogues to more general structures—is that in any one task, e.g., in discrimination learning, structures (or learning sets) are developed which are formally identical with those of the succeeding and preceding tasks. Thus, for example, the organism in choosing the "odd" member of a set of three objects engages at first in overt behavior which has the following characteristic: Take three objects, A, B, and C; the organism physically examines all three objects in some order, say A, B, C (i.e., he makes differentiating responses to them—Mandler, 1955—as he actually manipulates them); whenever the differentiating response to A is similar to the differentiating response to B (they are "alike") then a consummatory response will follow the choice of C (C is the odd object and its choice is reinforced); if the differentiating response to B is different from that made to A, then B will be followed by a consummatory response; if the differentiating response to A is followed by a consummatory response, no further behavior is elicited. In all cases the consummatory response is the stop rule. At this point we are still dealing with overt behavior during the early stages of learning. However, as the sequence of differentiating and overt responses is integrated, a structure of that sequence will be developed. This structure has three nodal points each of which represents a differentiating response and a choice response, none of which needs to be made overtly and the sequence can now run off in the following manner: Look at one object and make a differentiating response, look at the second object and, if the same differentiating response occurs as the one made to the first, then make an overt choice of the third object. Note that the matching requirement need not be held for more than two objects. Assume that A, B, and C have the characteristics k, k, and l, respectively. C therefore is the odd object. If the subject first examines A and then B, then the object following a match is the correct one; if he examines them in the order A, C, and B then a repetition of the sequence will lead to the same occurrence of a choice following a match (he has examined A, k; C, l; B, k; A, k; and now chooses C); if the order is C, A, and B, a return to C will lead to a similar result. In other words, the only required structures are an orderly sequence of covert choices and an overt choice or act following two similar differentiating responses. What the analogic structure permits the organism to do is to eliminate the overt examination of *all* the objects in order to arrive at the consummatory response.[3]

This exercise has been used to demonstrate *one way* in which cognitive categorical behavior may be slowly built up out of regular overt behavior sequences. One of the major variables of associative theories—frequency—appears to be an important antecedent of simple structures.

The development of analogic structures, and the resulting cognitive behavior is particularly relevant to verbal behavior which, in the natural setting, demonstrates high degrees of learning as well as cognitive correlates. And finally, while the apparent cognitive sequelae of overlearning have been demonstrated in mammals (monkeys, men, and rats), studies by Warren and his associates (Warren, 1960; Warren, Brookshire, Ball, & Reynolds, 1960) fail to find these effects in fish and chicks. More comparative data will have to be collected before we can speculate about the phylogenesis of cognition.

---

[3] The similarity between the process described here and the simulation of cognitive processes with computers (e.g., Newell & Simon, 1959) is obvious.

OVERLEARNING AND TRANSFER

In applying the notions developed above to the effects of overlearning on transfer of training, we shall proceed cautiously before invoking cognitive or structural explanations. We shall first review some of the typical findings in the literature, then outline the major variables apparently involved in the phenomena, and finally examine the extent to which associative factors may account for them, before invoking analogic structures as operative in these situations. The major independent variable with which we will be concerned is the degree of training on the original task, in particular when such training goes beyond mere mastery of the original task. We will be primarily interested in transfer situations which exemplify the Müller-Schumann paradigm, the A–B, A–C situation where identical or highly similar stimuli are used in both tasks but where the required response varies from the original to the transfer task. Overlearning beyond the attainment of asymptotic behavior provides repeated trials which are topographically highly similar from occasion to occasion, i.e., the same sequence of behavior occurs again and again. However, in some situations such as the learning of a complex maze by a rat or the learning of a paired-associate list by a human subject, it is difficult to specify when asymptote has been reached. In the case of the rat, some choice points will be learned very early and while the animal may still be making errors at some other choice points, he is at the same time overlearning the earlier acquisitions. In the case of verbal learning, e.g., of a list of 10 paired associates, the subject may be overlearning some pairs by the time he reaches even a low criterion of 3 or 4 correct out of 10. In this case, learning the list to a criterion of no errors will include some overlearning of some of the pairs in the list. Therefore, some data from the literature will be presented which are concerned with relatively lower degrees of training of the entire task but which probably include overlearning of some parts.

First, we can examine some typical results from the literature on the effects of degree of A–B training on A–C performance. The degree to which any one of these examples conforms to the ideal A–B, A–C situations does, of course, vary. Such variation extends from the ideal, where A is in fact an identical stimulus situation in both original and transfer tasks, to some degree of similarity between the original and transfer task. Similarity may imply the presence of common elements in the two situations or other constants across situations which are likely to elicit the original response.

Figure 1 shows some typical results from both animal and human studies. In each case the learning of the transfer situation with no prior training has been used to estimate the point of zero transfer and amount of transfer is plotted against number of trials on the original task. Performance for groups that have had some original learning

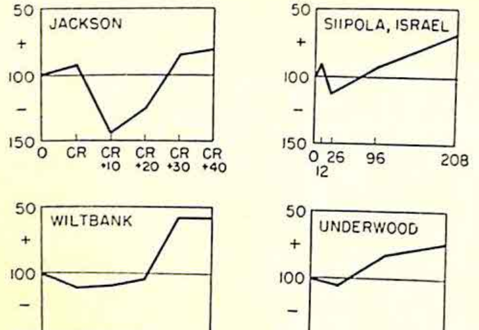


FIG. 1. Transfer of training as a function of degree of original learning. (See text for explanation.)

prior to the transfer task is represented in terms of percentages of the performance of the group with no prior training.

One of the most striking aspects of these graphs is the similarity of the general function of the relationship between transfer and degree of prior training. An initial negative transfer effect is followed by either a return to zero transfer, or the appearance of large positive transfer effects. The effect is **U** shaped in all instances.

The eight graphs in Figure 1 represent animal studies on the left and human studies on the right. Jackson (1932) trained his first group of rats to a criterion of 4 out of 5 correct trials on a multiple-cul maze. His second, third, fourth, and fifth group learned to criterion plus 10, 20, 30, and 40 overlearning trials, respectively. Transfer to a similar maze shows an initially large negative transfer effect followed by a slight positive effect with high degrees of overlearning. Wiltbank (1919), in a study using similar apparatus but lower degrees of learning (0, 2, 4, 8, 16 trials, and a mastery criterion), obtained highly similar results, though less negative and more positive transfer.

The next graph represents Reid's (1953) study, in which animals learned a black-white discrimination in a **Y** maze. They show negative transfer when reversed to the previously nonreinforced stimulus *if* they were trained only to criterion on the first task. However, with 50 overlearning trials we find less negative transfer, and positive transfer for the group that had 150 overlearning trials. Pubols (1956), in replicating this particular study, finds a similar effect though no positive transfer for the highly overlearned group. Galanter and Bush (1959) have presented some data on reversal learning which fail to show any positive transfer following high degrees (144 trials) of overlearning. Their animals, however, were run under conditions of distribution which resulted in much more spaced trials than Pubols used. Such a condition might be unfavorable to the development of stable structures.

The last graph for the animal studies represents a study by Bruner, Mandler, O'Dowd, and Wallach (1958). In this study animals were trained to learn a left-right alternation sequence in a four-unit maze. After varying degrees of acquisition in this original task, they were switched to the mirror image of the original task, e.g., the right-left-right-left response. Animals either learned to a criterion of 80% correct or were given 20 and 80 overlearning trials, respectively. Once again we find an initial negative transfer effect followed by a large positive transfer effect with 80 overlearning trials. The second (dotted) curve in this graph represents a group which was run when 36 hours hungry, the first group was run when 12 hours hungry.

We can now turn to the human studies. The first example is a study by Siipola and Israel (1933) in which subjects learned various telegraph codes to letters of the alphabet as stimuli. This study is not strictly an A–B, A–C situation. It involved the recombination of stimuli and responses in the transfer task. The same stimuli and same responses as in the original task were used in the transfer task with different combinations of stimuli and responses. Four groups were given different degrees of training; their means were 12, 26, 96, and 208 trials on the original task, respectively. The last two groups received high degrees of overlearning. In this particular case there is a slight positive transfer effect for the first group which is followed by a negative transfer effect and then by high positive effects for the last two groups.

Underwood (1949) used lists of 10 paired-associate adjectives. The most marked effect is the high degree of positive transfer with high degrees of overlearning, with only little negative transfer for the group receiving the least amount of training on the original task. In this particular situation, the first group had 3 correct out of 10 pairs, the second 8 correct out of 10 pairs, and the third group one complete correct trial plus five additional trials. The absence in this particular case of any large negative effect may be due to the lack of a warm-up control which we will discuss presently. Underwood also gives the number of errors that were made during the first two anticipation trials (Trials 2 and 3) of the transfer task. These are shown in the next graph where we find negative transfer effects for the least trained group which change to zero or positive transfer for the more highly trained group.

The next graph shows a study by Mandler and Heinemann (1956) in which subjects learned three-consonant nonsense syllables as responses to one digit numbers. The first overlearning group was given 10 trials beyond complete mastery of the list, the second 30 overlearning trials, and the third 50 overlearning trials. Again we find initial negative transfer effects followed by zero or positive transfer.

In examining these findings, five general processes will be considered: warm-up and learning set effects, response competition, stimulus discrimination, response learning, discontinuous structural factors.

Warm-up and learning set effects have been grouped together because—in the kinds of design presented here—they are usually confounded. Thus, it is often not determinable to what extent an animal's prior experience in a maze produces varying degrees of specific postural and attentive habits (warm-up) as against nonspecific structural effects (learning set). Presumably both kinds of effects are operating and may, for present purposes only, be considered together. But in order to evaluate the effects of overlearning on transfer, these more general experiential factors should be separated somehow from factors assignable to the competition and facilitation between original and transfer responses. We can examine the ideal model and determine what kind of control groups are necessary to separate situational and nonspecific transfer effects from those transfer effects which are peculiar to the juxtaposition of the particular original and transfer tasks. The model can be conceptualized as follows:

	Original task	Transfer task
Ideal paradigm	A–B	A–C
Control condition	D–E	A–C

The first line, A–B, A–C, is the ideal paradigm. However, any experience with the general class of tasks which A–C represents may result in transfer effects not due to A–B alone. In the second line the necessary control conditions are shown. In this case a D–E situation is necessary, which must be learned to the same degree as A–B was learned but which contains neither the stimuli of the A–C situation nor the responses of either the A–B or the A–C situation. In simple paired-associate learning this is not difficult to accomplish. Obviously we can have control groups which learn a series of paired-associate adjectives as in A–B, but with adjectives in the stimulus and response position which are different from either A, B, or C. A brief example of what happens when one does or does not use controls for this kind of situation is shown in Figure 2 which recapitulates the data from Mandler and Heinemann (1956) but incorporates a control for warm-up and learning-set effects. The first graph of the figure shows the raw data for the A–B, A–C situation, and in the second graph is shown the raw data for a D–E, A–C situation. (For comparison purposes the large positive effects of an A–B, C–B condition are also shown.) In both cases we find increasing positive transfer as a function of overlearning on the original task. However, the positive transfer for the control condition is greater for the first two groups and less so for the last group. The resultant is shown in the third graph, which shows initial negative transfer followed by positive transfer.

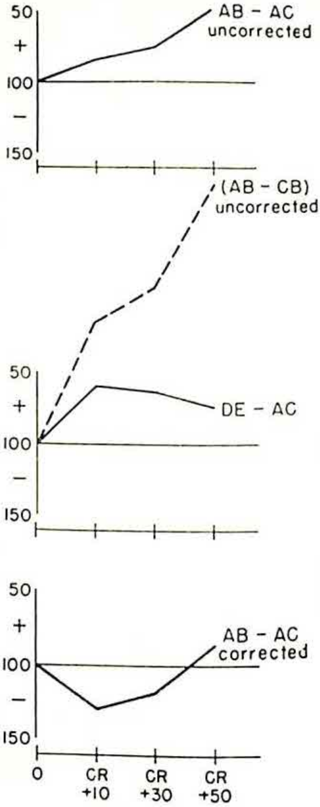


FIG. 2. Effect of warm-up correction on transfer effects. (See text for explanation.)

The paired-associate situation is relatively easy to consider in controlling for warm-up and learning to learn effects. It becomes more difficult when we think about some of the animal studies. It may be difficult to specify a warm-up condition which will separate practice on a specific maze from general maze or situational experience, but such control groups seem to be essential.

Some of the studies do suggest possible control groups. In Reid's (1953) study, for example, it might be possible to use control groups who learn to discriminate but do not learn a differential black-white response. Reid argues for the acquisition of a new response, the response of discriminating, which does not occur until late in overlearning. If animals were to learn a discrimination between a triangular and a circular stimulus and given various degrees of training on this discrimination before being transferred to a black-white discrimination, we would then have a control group which does not have the black-white practice, but does have practice in discriminating. Any resultant advantage which still accrues to the group that had its prior practice on the black-white discrimination would be specific to that particular situation and not related merely to being in a discrimination task. In general there tends to be some confusion between specific transfer effects and general transfer effects due to warm-up or learning to learn. Transfer studies must find some way of separating general from more specific transfer effects.

The first of the specific transfer effects to be discussed is the classical one of response competition or interference. The response which was appropriate in the original task is elicited in the transfer task, particularly if the stimulus situation in the two tasks is highly similar. This effect has led past investigators to call the A–B, A–C paradigm the negative transfer paradigm par excellence. Despite the failure to find such a result in the overlearning studies, the available evidence in transfer studies suggests that the interference or competition hypothesis need not be abandoned. What appears to be the case is that interference effects occur very early during the transfer task and that "factors making for positive transfer . . . do not gain the ascendancy over the negative tendencies until the latter stage of the learning process" (Wiltbank, 1919, p. 42).

The important point is that the probability that Stimulus A, which is present in both situations, will elicit Response B, which was learned in the original task and which interferes with the performance of Response C required in the transfer task, does increase as a function of prior A–B pairings. But, despite this increase in intrusion errors, the overall efficiency in learning the transfer task is increased.

Underwood's (1949) data show large interference effects during the first two anticipation trials, probably primarily due to intrusions of the now erroneous response during early learning of the transfer task. In animal studies, Reid (1953) has shown that his highly overtrained animals make many more intrusion errors, i.e., going to the now incorrect stimulus, than do the less trained animals. The Bruner et al. (1958) study also indicates that the amplitude (in time) of the now incorrect "left" response at the first choice point of the transfer task increases as a function of degree of original training—though in the 12-hour deprived group only.

As far as stimulus differentiation is concerned, the more experience the subject has had in differentiating the stimuli used in transfer, the less likely we are to find intratask competition among various stimuli (Gibson, 1940, 1953). But whenever the stimulus situation in the two tasks is identical, as in the A–B, A–C paradigm, stimulus differentiation is only relevant to the degree that the subject has learned a consistent response to the stimulus in the original task—nothing new needs to be learned about the stimulus in the transfer task.

More important than stimulus learning is response learning. The degree to which the organism "knows the response" is an important factor in his performance of the transfer task. If that task requires a new response which has not been previously practiced and integrated, the early trials of the transfer task will be devoted to response learning as well as to the elimination of competition with the responses of the original task.[4] Special attention must be paid here to extra-experimental experience. Responses may well have been learned prior to the transfer task in situations other than the original task. For example, in paired-associate learning with adjectives it is obvious that whatever adjectives are used in the transfer task, the subject has had some prior experience in learning how to make the responses—a condition not applicable to some nonsense syllable learning experiments where the subject has to learn the response *de novo*.

Before considering discontinuous cognitive effects, we shall summarize the hypothesized effects of the other four factors on the transfer task as a function of degree of training on the original task.

1. We may tentatively assume that warm-up activity and learning sets both have a monotonic facilitating effect on subsequent behavior.

2. Degree of prior learning of an "incompatible response" has a monotonically increasing interfering effect. It also appears to be the case that these effects, while manifested in intrusion errors, do not necessarily hinder the rapid acquisition of the new response.

3. Stimulus discrimination also seems to have a monotonic facilitating effect on subsequent learning.

4. Finally, response learning facilitates subsequent acquisition of a new response to the extent to which the required response has been previously practised.

There seems to be some evidence that Bair's (1902) and Wiltbank's (1919) conclusion that negative effects (response competition in our terms) are maximal during early learning and that positive effects do not appear until the later stages of learning is essentially correct. What may also be the case is that during the later stages of acquisition or performance new responses are learned which can be transferred to the new task. The "observing response" or "response of discriminating" described by Wyckoff (1952), Reid (1953), Pubols (1956), and others, represents the learning of a new response which is easily transferred to the new situation and which, as Pubols suggests, is acquired during the later stages of the learning process. However, a recent study by Birch, Ison, and Sperling (1960) using single stimulus presentations showed essentially similar results for the effects of overlearning on discrimination reversal. They conclude that their findings "effectively rule out interpretations . . . in terms of responses of discriminating simultaneously presented discriminanda" (p. 40). Birch et al. also appear to dispose of the suggestion by D'Amato and Jagoda (1960) that "overlearning should not facilitate reversal learning in a successive discrimination situation, since avoidance of S_- is maintained by the very nature of the situation" (p. 259).

If we apply these various factors to the studies summarized in Figure 1, one possible conclusion is apparent. If negative factors such as response competition are dominant after low degrees of original learning, and if positive factors such as warm-up and learning set do not appear in full force until later stages of learning, i.e., during overlearning, then an initial negative transfer effect followed by a reversal in a positive direction, after the organism has had longer experience with the original task, is predictable. Such an explanation need not invoke any cognitive discontinuous processes. Wiltbank's, Jackson's, Reid's, and Pubols' studies all might be explained in terms of specific learning occurring during the original task. The experiment by Bruner et al. (1958) designed specifically to demonstrate the acquisition and use of rules, is probably also a case of response transfer, since an integrated right-left-right-left sequence could easily be transferred to a left-right-left-right sequence—large chunks of the required behavior are practiced during original learning and probably well integrated during overlearning of the original task. A more crucial question would be raised if a control group similar to the one suggested for the Reid study were included. If animals were trained on a circle-triangle discrimination and then transferred to a black-white discrimination,, their behavior on the black-white task would provide a baseline for evaluating reversal effects. Thus, once reversal data are adjusted for the effects of varying degrees of prior maze experience, it will be possible to state whether the overlearning reversal effects can be explained in terms of maze-specific behavior or whether animals during overlearning acquire something other than warm-up, learning sets or discriminating responses, namely some cognitive structure about black and white stimuli.

---

[4] Underwood and Schulz (1960) have elaborated earlier suggestions concerning the importance of this variable in verbal learning.

In the situation where the subject learns to make a new response to an old stimulus, the mere learning of the old response is irrelevant. If we further assume that stimulus discrimination is not a major factor contributing to transfer, we are left with three effects. Response competition, the major negative effect, should increase to asymptotic levels as a function of prior pairings of old stimulus and old response. Warm-up and learning-set effects should increase more markedly during the later (overlearning) trials and produce positive transfer effects. The learning of new nonspecific responses (e.g., observing responses) probably also appears late during original learning and has positive effects. However, once the appropriate control groups are used, all of the positive effects should be controlled and we should be left only with the negative effect of response competition. When these negative effects change to zero or positive effects despite the use of the appropriate controls (as in Mandler, 1954b; Mandler & Heinemann, 1956; and Siipola & Israel, 1933; and possibly in Underwood, 1949) it is appropriate to invoke some other hypothesis. We have suggested that overlearning experience with the old response and the formation of an analogic structure permits the subject to manipulate that response, to "think about" the problem without making overt errors. Just as we assume that the animal in the oddity problem can use cognitive structures to eliminate the need for overt choices, and in discrimination problems can manipulate black-white choices covertly by inhibiting responses to the "incorrect" stimulus, the subject in the A–B, A–C situation may be able to eliminate overt B responses.

Starting off with the assumption of a continuity from simple associative structures to cognitive structures, we have examined one area of research in detail to find evidence for the appearance of cognitive structures. Our investigation has uncovered a variety of associative factors which are adequate to explain what may at first glance appear to be discontinuous behavior. This will, of course, be no surprise to the committed S–R theorists who have always maintained that most, if not all, observations on apparently cognitive, categorical behavior will, in the long run, be amenable to an associationistic explanation. While we would agree that sophisticated associationist positions (e.g., Miller, 1959; Spence, 1956) have been too readily abandoned by their critics, we would maintain that some behavioral phenomena defy current associationistic interpretation. However, recourse to cognitive structures should await the verdict of the laboratory; the proper use of some of the control conditions suggested above and a more fine-grained analysis of warm-up effects and learning sets will determine which of the phenomena discussed here will ultimately call for cognitive explanations.

SUMMARY

Given the distinction that associationist and cognitive theories make in respect to the definition of response, we have examined the proposition that one source of cognitive behavior may be found in associative antecedents. The argument for a transition from associations to cognitions was buttressed by reference to the development of learning sets and developed out of the notion of response integration and the emergence of analogic structures. Empirical evidence on the effect of overlearning on transfer of training was examined in detail to differentiate findings that seem to be amenable to associative explanations from those that seem to call for structural, cognitive explanations.

REFERENCES

BAIR, J. H. The practice curve: A study in the formation of habits. *Psychol. Rev. Monogr. Suppl.*, 1902, **5** (2, Whole No. 19).

BIRCH, D., ISON, J. R., & SPERLING, SALLY E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

BRUNER, J. S., MANDLER, JEAN M., O'DOWD, D., & WALLACH, M. A. The role of overlearning and drive level in reversal learning. *J. comp. physiol. Psychol.*, 1958, **51**, 607–613.

CAMPBELL, D. T. Operational delineation of "what is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167–174.

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials on discrimination reversal. *J. exp. Psychol.*, 1960, **59**, 254–259.

DUNCAN, C. P. Description of learning to learn in human subjects. *Amer. J. Psychol.*, 1960, **73**, 108–114.

GALANTER, E., & BUSH, R. R. Some T-maze experiments. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Pp. 265–289.

GIBSON, ELEANOR J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GIBSON, ELEANOR J. Improvement in perceptual judgments as a function of controlled practice and training. *Psychol. Bull.*, 1953, **50**, 401–431.

HAMILTON, C. E. The relationship between length of interval separating two learning tasks and performance on the second task. *J. exp. Psychol.*, 1950, **40**, 613–621.

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

IRION, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55**, 336–341.

JACKSON, T. A. General factors in transfer of training in the white rat. *Genet. psychol. Monogr.*, 1932, **11**, 1–57.

KENDLER, H. H. "What is learned?" A theoretical blind alley. *Psychol. Rev.*, 1952, **59**, 269–277.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244. (a)

MANDLER, G. Transfer of training as a function of degree of response overlearning. *J. exp. Psychol.*, 1954, **47**, 411–417. (b)

MANDLER, G. Perceptual learning: Stimulus differentiation or differentiating response? *Psychol. Rep.*, 1955, **1**, 79–82.

MANDLER, G. The warm-up effect: Some further evidence on temporal and task factors. *J. gen. Psychol.*, 1956, **55**, 3–8.

MANDLER, G., & HEINEMANN, SHIRLEY H. Effect of overlearning of a verbal response on transfer of training. *J. exp. Psychol.*, 1956, **51**, 39–46.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

NEWELL, A., & SIMON, H. A. *The simulation of human thought.* Santa Monica, Calif.: Rand, 1959.

PIAGET, J. *The origins of intelligence in children.* New York: International Univer. Press, 1952.

PUBOLS, B. H. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

ROZEBOOM, W. W. "What is learned?" An empirical enigma. *Psychol. Rev.*, 1958, **65**, 22–33.

SIIPOLA, ELSA M., & ISRAEL, H. E. Habit interference as dependent upon stage of training. *Amer. J. Psychol.*, 1933, **45**, 205–227.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

THUNE, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40**, 423–438.

TOLMAN, E. C. Sign-gestalt or conditioned reflex. *Psychol. Rev.*, 1933, **40**, 246–255.

UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WARREN, J. M. Reversal learning by paradise fish (*macropodus opercularis*). *J. comp. physiol. Psychol.*, 1960, **53**, 376–378.

WARREN, J. M., BROOKSHIRE, K. H., BALL, G. G., & REYNOLDS, D. V. Reversal learning by white leghorn chicks. *J. comp. physiol. Psychol.*, 1960, **53**, 371–375.

WILTBANK, R. T. Transfer of training in white rats upon various series of mazes. *Behav. Monogr.*, 1919, **4**, No. 17.

WYCKOFF, L. B. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received February 15, 1961)

Psychological Review
1962, Vol. 69, No. 5, 428–449

STIMULUS–INTENSITY EFFECTS IN RESPONSE EVOCATION[1]

R. A. CHAMPION[2]

University of Sydney

It is a well-established fact that the strength of a response varies directly with the intensity of its eliciting stimulus. The relationship holds for a wide variety of reactions and within the range of stimulus intensities from the lower absolute threshold to the level at which the receptors suffer physical damage. Despite the widespread recognition of the phenomenon, two basic problems remain unsolved. First, it is not known exactly what property of the stimulus situation causes the effect, i.e., there is no precisely articulated theory of stimulus-intensity effects which is firmly established. Second, the question arises as to whether the amount of true *learning*, as distinct from performance, varies with the intensity of the eliciting stimulus during training. Beginning with an outline of the two theories advanced to explain stimulus-intensity effects, the following discussion deals briefly in that context with attempts to assess the role of stimulus intensity in learning; the greater part of the paper is then taken up with a consideration of some empirical evidence on response evocation which seems to allow a choice between the two theories.

THEORIES OF STIMULUS-INTENSITY EFFECTS

The first theorist to deal with the stimulus-intensity problem in formal terms was Hull (1951). In typical fashion he began by making it the subject of a postulate and, in effect, assumed a constant relationship between response strength and stimulus strength, other things being equal. He named the effect "stimulus-intensity dynamism" and represented it in his behavior system with the intervening variable V. "Dynamism" then became the stimulus-intensity component of reaction potential, acting multiplicatively with the other positive factors in the system. Hull (1952) later proceeded beyond the mere assumption of a relationship to a more elaborate theory in which he attempted to explain the effect with the concept of the molar stimulus trace. According to this formulation, the physiological effect of stimulation, occurring when a brief stimulus impinges upon a suitable receptor, reaches a maximum about 450 milliseconds after the beginning of the stimulus. Hull thereby limited the stimulus-dynamism process to the *onset* of a stimulus and to short-latency responses such as the eyeblink. He also gave the appearance of allowing for the effect of stimulus intensity on amount learned by distinguishing between V_1, the dynamism of the stimulus during the *learning* of the response, and V_2, the dynamism of

---

[1] Part of the apparatus used in these studies was provided by Commonwealth of Australia Research Grant No. 2002.

[2] The author is indebted to P. G. Ward for assistance in interpreting the data of Experiment II and to Joan E. Jones for assistance in gathering and treating the data of Experiments III and IV.

the stimulus during the later *evocation* of the response. At one point Hull (1951) did suppose that *habit strength* during learning depended on V_1, but he later (1952) shifted to the idea that *reaction potential* during training involved V_1 as a component. It would appear that, in his later theorizing, Hull confined V to the strength of the signaling stimulus trace as dependent on the age of the trace instead of on the intensity of the stimulus. Thus he implied that V_1 affected actual learning through stimulus-response asynchronism rather than through the strength of the eliciting stimulus. Hull was not clear on this point, but the final version of his theory does not seem to allow the deduction that amount learned increases with the intensity of the actual stimulus.

The only other theory of stimulus-intensity effects has been proposed by Perkins (1953) and Logan (1954). Working independently these contributors arrived at theories which are so similar that they will be treated as one for the present purpose. This theory is based on the assumption that, as a given response is being elicited by a given stimulus with reinforcement, so the stimulus gains excitatory properties which generalize to similar stimuli; conversely, the theory assumes that since the occurrence of the same response to other stimuli in the situation is not reinforced, then these other stimuli gain inhibitory properties which also generalize. The latter class of stimuli is said to be made up of "background cues" or intertrial stimuli, or to constitute the "contextual environment." This state of affairs, the result of differential conditioning, is represented schematically in Figure 1. If a tone is used as a conditioned stimulus (CS) in a classical conditioning situation, for example, then it is assumed to acquire the same degree of gross excitatory potential in separate groups trained with a weak tone (W in Case A on the left) and a strong tone (S in Case B on the right), and this generalizes along the stimulus-intensity dimension as shown. When a response is given to the intertrial stimulus of "silence" (0, the ambient noise level) during this training period, it is not reinforced and the intertrial stimulus acquires inhibitory potential which also generalizes. The further assumption that inhibition is to be subtracted from gross excitation to

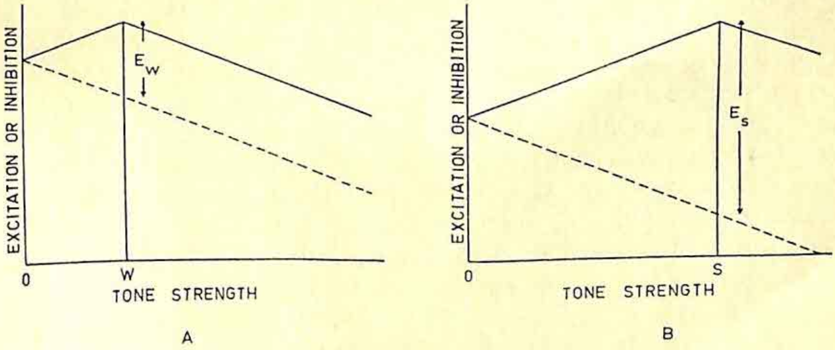


FIG. 1. Schematic diagram of the hypothetical gradients of excitation (continuous line) and inhibition (broken line) after reinforcement of responses to the onset of a weak tone (W) as CS (Case A) and a strong tone (S) as CS (Case B) with nonreinforcement of responses to the intertrial stimulus of "silence" (O). (Net excitation is greater for the strong CS—E_s—than for the weak CS—E_W.)

give net excitation allows the Perkins-Logan theory to predict greater net excitation for a strong tone (E_S) than for a weak tone (E_W), with the consequent expectation of a stronger response in the former case.

The Perkins-Logan theory assumes that amount learned is independent of the intensity of the eliciting stimulus. If it were to be shown that learning is superior with a stronger CS then the case for the deduction made above would be strengthened, gross excitation being greater with the strong than with the weak tone. A later deduction taking drive level into account, however, would be altered in this case. The present application of the theory also assumes that the intertrial or "background" stimulus acquires just enough inhibitory potential to prevent a response occurring between reinforced presentations of the stimulus; this should be the case, because as enough inhibition to block the response is generated so the response will no longer occur unreinforced and no more inhibition can accrue to the stimulus. It will also be noted that the generalization gradients have been depicted as linear, following Perkins (1953). This has been done in the present instance because it is felt that the exact shape is yet to be firmly established. Perhaps the weight of evidence favors the existence of negatively accelerated gradients (e.g., Raben, 1949) and one of the following deductions is made on this basis, but the remainder of the argument is not significantly affected by the assumption of a variety of other types of gradient.

Further consideration of the account of the Perkins-Logan theory given above, read in conjunction with Figure 1, may lead to the inference that stimulus-intensity effects are to be understood as the result of discrimination learning. In terms of the example depicted in Figure 1, a strong stimulus produces a stronger response not as a result of any inherent dynamism but rather because its presence is more readily discriminated from its absence. This parallel may be drawn in certain instances and will be considered wherever possible in the following discussion, but it is an inexact analogy in some respects and is therefore potentially misleading. Basically, the Perkins-Logan interpretation of the stimulus-intensity situation is put in terms of differential conditioning rather than traditional discrimination learning; i.e., the interest lies not in which of two or more responses will occur to a given stimulus, nor in which of two or more stimuli will elicit a given response, but in how strongly a given response will occur when elicited by a given stimulus. To put the matter another way, although the effect on a given response of variations in the strength of its eliciting stimulus depends upon generalized inhibition from the intertrial or background stimulus, these two stimuli are not necessarily presented simultaneously and no discrimination is therefore required. In the example cited above there is temporal or successive contrast between the positive CS and the negative intertrial stimuli. The case of a rat learning to run to a black card on a white background, with spatial or simultaneous contrast of "cue" and background stimuli, to which the Perkins-Logan theory has also been applied (e.g., Nygaard, 1958), resembles true discrimination learning more closely. It is still a form of spatial learning, however, for the cue and background stimuli hold constant positions in space and are not reversed in the one learning series for a given animal.

A final point of interest in the exposition of the differential conditioning theory is the probability that it applies to variations in the *quality* as well as in the intensity of the stimulus. Predictable effects should ensue whenever the eliciting and background stimuli lie on some common qualitative dimension, with intensity held constant. There is no reason why experiments should not be conducted to test for stimulus-quality effects, but with the possibility considered, the present discussion will be limited to intensity effects.

STIMULUS INTENSITY AND LEARNING

The second of the two problems stated at the outset concerned the possibility that a response is better learned if the eliciting stimulus is strong during learning. A number of attempts have been made to investigate this possibility, but no satisfactory test has yet been devised, so that the problem remains unsolved and a challenge to future workers. For this reason the problem cannot be pursued at length now, although a brief account of its nature may clarify the discussion which follows.

The latest attack on the problem formed part of a study of eyelid conditioning conducted by Walker (1960), which may be used for illustrative purposes. Following the lead of Grant and Schneider (1948, 1949), Walker used a factorial design; she trained one group of subjects with a strong CS and another group with a weak CS, and then shifted half of each of these groups to weak and strong CSs, respectively, in extinction, leaving the other half of each group on the CS strength used in training. The nature of this experimental design is represented in Table 1, where the first letter in each cell stands for CS intensity during training, the second letter stands for CS intensity during extinction, and the measures are taken on the extinction series. Walker sought to determine the effect of CS intensity on habit strength by comparing the performance in extinction of the two training groups, represented in the two rows of Table 1, but there proved to be no significant difference.

TABLE 1

THE FACTORIAL DESIGN USED BY WALKER (1960)

Training CS intensity	Extinction CS intensity	
	Strong	Weak
Strong	SS	SW
Weak	WS	WW

Unfortunately, this aspect of Walker's (1960) experiment is rendered invalid by the possible interfering effects of the shifts in CS intensity from training to extinction which occurred in half the subjects. The impact of these shifts may be treated in terms of both the Hullian and Perkins-Logan theories. Taking Hull's theory first, the shifts should bring changes both in generalized habit strength due to stimulus-intensity generalization and in stimulus-intensity dynamism (from V_1 to V_2). The former effect would always cause a loss of response strength whereas the latter factor could produce either an increase or a decrease therein. These two factors might obscure any effect of V_1 on ${}_sH_R$ and until some determination is made of their differential weights in contributing to ${}_sE_R$ no allowance can be made for them. It is thus impossible to infer from Walker's data that amount learned is independent of CS intensity. Walker did not make this inference; she suggested instead that extinction data may not provide a proper basis for testing the hypothesis but failed to give any reason for the suggestion.

Further reference to Figure 1 will

show how the Perkins-Logan theory also leads one to believe that shifts in CS intensity after training might obscure any effect on amount learned. On the basis of this theory the expected performance of Walker's (1960) four groups in extinction (Table 1) may be set down thus: WW—a continuous decline due to inhibition accruing to the CS as well as to the intertrial stimuli; WS—as for WW; SS—a continuous decline from a higher level of performance than that initially present in WW and WS; SW—a sharp decrease with the shift, from the same higher level, followed by a slower decline. If these differential effects are to be expected with a null hypothesis that CS intensity does not affect learning then it is clear that no alternative hypothesis may be subjected to a proper test. Recalling that Walker obtained no significant effects in extinction, with unconditioned stimulus (UCS) intensity held constant, it might seem that the predictions derived from the Perkins-Logan theory were not verified. It is doubtful, however, that Walker's analysis of the data did justice to the predictions, no account being taken, for example, of trends over successive trials in extinction. On the other hand it is conceivable that the discrepancy is due to the occurrence of superior learning with a stronger CS.

To return briefly to the analogy with discrimination learning, the shifts in CS intensity from training to test which are used in this type of study are reminiscent of the "transposition phenomenon." The analogy is still inexact, for the original form of transposition required the preservation of a relationship between two stimuli when the absolute intensity of both was changed; argument then took place as to whether the relationship was vital to the learner (gestalt theory) or to the theorist (S-R theory). In the factorial studies of stimulus-intensity effects, the absolute value of only the CS or cue stimulus is changed and that of the intertrial or background stimuli remains constant. Such experiments could be conducted, of course, with shifts in the intensity of both sources of stimulation, so as to approach true transposition more closely.

This short account of the possible role of stimulus intensity in learning, with Walker's (1960) study as an example of relevant experimental evidence, should suffice to lend weight to the claim that no satisfactory test of this aspect of the stimulus-intensity phenomenon has yet been devised. It is therefore fortunate that a clear distinction may be made between effects on learning and on performance. The remainder of this paper is accordingly devoted exclusively to effects on performance; it is hoped that the preceding examination of effects on learning will allow the reader to bear this important distinction in mind.

STIMULUS-OFFSET EFFECTS ON PERFORMANCE

In considering the implications of his theory, Logan (1954) suggested that the offset of a stimulus should be as effective as its onset in eliciting a response and that the strength of the response should vary with the intensity of the offset stimulus. It is difficult to see how a strict interpretation of Hull's stimulus-dynamism principle could predict such an outcome. Hansche and Grant (1960) have shown that offset and onset are equally effective as a CS in eyelid conditioning, and Schwartz (1958) found avoidance response strength in rats to vary directly with degree of change in CS intensity, regardless of

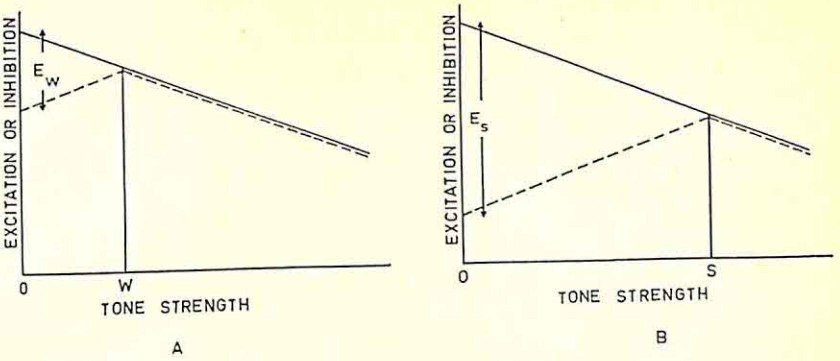


FIG. 2. Schematic diagram of the hypothetical gradients of excitation (continuous line) and inhibition (broken line) after reinforcement of responses to the offset of a tone (onset of "silence") as a CS with nonreinforcement of responses to a weak tone (Case A) or a strong tone (Case B) as the intertrial stimulus. (Net excitation is greater with the offset of the strong tone—E_s—than with the offset of the weak tone—E_W.)

the direction of the change. The detailed application of the Perkins-Logan theory to the case of stimulus offset as a CS is represented schematically in Figure 2, in which the symbols are the same as for Figure 1. In this situation it is the state of "silence" which should acquire excitatory properties, whereas the weak or strong intertrial stimuli come to inhibit the response. Even so, the net excitation with a strong intertrial stimulus (E_S) is still greater than that with a weak intertrial stimulus (E_W) after the excitation and inhibition have been built up in the course of training. Before any differential conditioning, on the other hand, there should be no difference in the strength of the responses elicited by the two degrees of offset. An experiment was conducted to test these deductions; it allowed the findings of Hansche and Grant, and of Schwartz, to be confirmed, but made wider generality of application possible by the use of the long-latency GSR.

EXPERIMENT I

The subjects were 24 male and 24 female volunteers from an introductory course in psychology at the University of Sydney. The intertrial stimulus was a 2000-cps sine-wave tone delivered to the subject through headphones at sound pressure levels of 80 db (strong) or 60 db (weak) and the CS was the offset of the tone for 2000 milliseconds. An attempt was made to choose sound levels such that the loud tone would not be so intense as to disturb the subject when presented almost continuously for 20 minutes and that the soft tone would not be so weak as to be below threshold for any subject. The UCS was a 90-cps spike-wave shock of 500-millisecond duration which was presented 500 milliseconds after the offset of the intertrial stimulus as CS. It was set at 2.27 milliamperes as measured with a 15,000-ohm resistor in place of the subject and was delivered through thimble-type electrodes of German silver .5 inch in length, attached to the third and fifth fingers of the right hand. The GSR was picked up with electrodes of a similar type 1 inch in length, attached to the second and fourth fingers of the right hand, through which was passed a constant direct current of 100 microamperes; the response was measured with a bridge-type circuit linked to an amplifier and ink-writing recorder. Conductive paste was used in conjunction with the shock and GSR electrodes. The duration of the CS, the UCS, and the CS-UCS interval was controlled by electronic timers. The subject was placed in a sound-reduced room completely separate from that containing the experimenter and the main apparatus.

The measure of performance used was the amplitude of the GSR on test trials (tone offset alone) which were interspersed with the training trials (tone offset followed by shock). Each series began with a test trial followed by one presentation of the UCS alone and a second test trial. Then three further test trials were given after 1, 2, or 3 training trials (third test trial), after 4, 5, or 6 training trials (fourth test trial), and after 7, 8, or 9 training trials (fifth test trial). Within this arrangement six orders of trials were used, in which test trials were presented after the following numbers of training trials: 1, 5, 9; 1, 6, 8; 2, 4, 9; 2, 6, 7; 3, 4, 8; 3, 5, 7. Thus the mean number of preceding training trials for Test Trials 1 through 5 were 0, 0, 2, 5, and 8 trials successively. The subjects were trained either with the offset of the strong tone as the CS (Group S) or with the offset of the weak tone as the CS (Group W). They were allocated to the two groups of the experiment and to the six orders of trials in a counterbalanced sequence imposed upon the order in which they appeared at the laboratory with the restriction that there be 12 men and 12 women in each of the two groups. The intertrial interval was at least 60 seconds and depended in part on the time taken to achieve a steady level of skin resistance; it rarely exceeded 75 seconds and averaged about 65 seconds.

The performance curves derived from the data are shown in Figure 3, in which the first point on the two curves is based on each subject's mean response over the first and second test trials. The performance of Group S was compared with that of Group W before any training trials and a *U* test showed that the difference was not significant ($z = 1.51$, $p = .13$). The nonparametric test was preferred because of some evidence of nonnormality in the distribution of scores. To ensure that the groups were comparable at this stage a *U* test was also applied to the responses on the second trial, consisting of the UCS (shock) alone. The mean values were 1.57 and 1.74 micromhos for Groups S and W, respectively, and the difference was not significant ($z = 0.22$, $p = .41$). Inspection of the results showed that Group S improved more than Group W in conditioning and the significance of this finding was tested by taking the difference between the amplitude of the response on the second and third test trials for each subject. A *U* test showed that the gain was significantly greater for Group S ($z = 2.12$, $p = .03$).

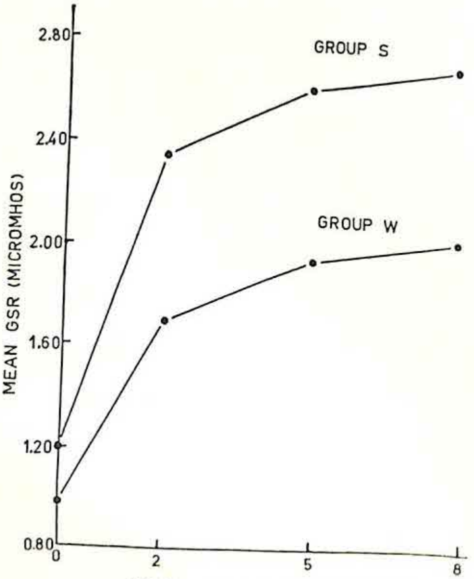


FIG. 3. Response strength on test trials in Experiment I as a function of the mean number of preceding training trials for Groups S (strong tone off) and W (weak tone off).

The results of this experiment confirmed that the offset of a stimulus may be used as a CS and that, in the course of acquisition, the offset of a strong CS comes to elicit a stronger response than the offset of a weak CS, as the Perkins-Logan theory predicts. It is emphasized that this effect is being treated as a matter of *performance* only (Figure 2), it is not being inferred that true *learning* is greater with the offset of the stronger CS. Hull's principle of stimulus-intensity dynamism was couched purely in terms of the absolute strength of the eliciting stimulus, acting through its onset. In formulating the principle, he cited evidence that rats jump faster to white than to black cards, but a law of absolute stimulus strengths is manifestly inapplicable to the case where animals respond more strongly to a black (low intensity) card on a white background than to a white (high intensity) card on a white background (Nygaard, 1958). It is exactly an effect such as this, on the contrary, which is predicted by the Perkins-Logan theory. A further difficulty in the application of Hull's theory to the present data is its limitation to short-latency responses and its failure therefore to explain such effects as have been obtained with the GSR. The two main results of Experiment I were the demonstration that "silence" is comparable in effectiveness with an audible tone, as a CS, and that the response elicited by a brief period of "silence" is stronger with a loud tone as the intertrial stimulus than with a weak tone between trials. The former result is comparable with the common finding in discrimination learning studies that, with a white card and a black card presented simultaneously, for example, either may be made the "positive" cue in the sense that approach to it is followed by reinforcement, provided allowance is made for initial position preferences. The second result of Experiment I is presumably analogous with the equally common finding that rate of discrimination learning is directly proportional to the degree of difference between the positive and negative cues, for example, that learning proceeds more quickly with white and black cards as the cues than with white and gray or with black and gray.

The results of Experiment I suggest a further prediction about generalization effects which follows the application of the Perkins-Logan theory to the stimulus-offset situation. It is first necessary, however, to deal with these effects as they occur in the more usual case of stimulus onset (Figure 1). The clearest evidence on stimulus-intensity generalization has probably been provided by Grice and Saltz (1950), who found asymmetrical gradients falling rapidly with test stimuli weaker than the training stimulus and more slowly for stronger test stimuli. Examination of Figure 1 shows that, according to the present formulation of the Perkins-Logan theory, tests with stronger stimuli should not yield a weaker response, as net excitatory potential stays constant. The reason why this deduction does not correspond with the facts may well lie in the assumption of linear gradients of generalized excitation and inhibition. If negatively accelerated gradients are substituted for the linear gradients of Figure 1 then tests with stronger stimuli should also produce a weaker response. The gradients of *net* excitatory potential should still be asymmetrical; Hull (1949) predicted this asymmetry by allowing for stimulus dynamism, but the substitution of negatively accelerated for linear gradients allows the Perkins-Logan theory to predict the same outcome. According to the latter theory, however, when the eliciting and background stimuli do not lie on any common dimension, then the gradients of *net* excitatory potential should be symmetrical about the value of the eliciting stimulus. This lack of communality and the consequent symmetry are more likely with qualitative differences than with changes in stimulus intensity.

The application of these considerations to the case of stimulus offset gives rise to the prediction which is still untested. When the ambient noise level ("silence") of, say, 50 db is used as the CS then tests of stimulus generalization may be made by briefly changing the intensity of the intertrial stimulus to some value other than 50 db. For example, the intertrial stimulus could be switched from 80 db to 90 db or to 70 db for 2000 milliseconds (see Case B of Figure 2). *With negatively accelerated gradients assumed*, a momentary increase or decrease in the intensity of the background stimulus should produce a response which is always weaker than the original CR but relatively stronger with the decrease than with the increase. Momentary decreases in the intensity of the intertrial stimulus *below* the intensity of the CS (50 db), however, should produce another gradient of net excitation. Thus a plot of the curve relating net excitatory potential to stimulus intensity should show two points of inflection, one at the value of the CS and another at the value of the intertrial stimulus. This effect would be easier to detect if the CS as well as the intertrial stimulus was set at some intensity above the ambient noise level, at 60 db, for example. Inasmuch as the more usual stimulus-onset situation involves an intertrial stimulus of some intensity, however weak, then the same effect should be found there, but its presence would be more difficult to detect for technical reasons.

MODIFIED HULLIAN THEORY

The results of Experiment I showed that a strict, literal interpretation of Hull's principle of stimulus-intensity dynamism must be abandoned, but they also suggest a modification by which Hull's approach to the problem might be sustained. Along with all other S-R theorists Hull assumed that a stimulus acts through its occurrence or presence and this is the source of difficulty in dealing with the cessation or "absence" of a stimulus as an active factor. The difficulty could be overcome with the supposition that a stimulus acts through any *change* in the intensity of the physical energies impinging upon the organism's receptors and that the effectiveness of the stimulation varies with the degree of change, *regardless of its direction*. An assumption of this kind is supported by physiological evidence for the existence of nerve fibers which respond only to changes in stimulation (e.g., Granit, 1947). Thus it is possible to contrast a neo-Hullian or modified stimulus-intensity dynamism principle with the differential conditioning theory of Perkins and Logan, and a second experiment was designed with this aim in view. In its revised form the stimulus-intensity dynamism principle should ascribe equal effectiveness to the onset and the offset of a stimulus, provided that degree of change is held constant, and should not lead one to expect any effect with a shift from onset to offset (or vice versa) as the stimulus over a series of trials with the same subject. According to the differential conditioning theory, on the other hand, once a given stimulus has acquired inhibitory properties for a given subject then negative transfer should occur if the same stimulus is then used to elicit the response in the same subject, as happens with a shift from onset to offset or vice versa. In order that due allowance might be made for the limitation of the Hullian formulation to short-latency responses, this experiment was conducted with a key pressing movement instead of the GSR.

EXPERIMENT II

The subjects were 12 members of undergraduate courses at the University of Sydney. The form of physical energy manipulated as an eliciting stimulus was the same tone as in Experiment I, again delivered through headphones at 80 db or 60 db SPL. The subject was placed in the sound-reduced room and asked to respond by pressing a telegraph key fitted with a small platform upon which rested the fingers of the right hand. Reaction times were measured with an electronic chronoscope. Each subject was presented with the four combinations of stimulus conditions in blocks of nine trials, viz., weak tone offset (A), strong tone offset (B), weak tone onset (C), and strong tone onset (D). With subjects numbered in order of testing, they underwent the four conditions in one session in the following orders: Subjects 1, 5, and 9—CDAB; Subjects 2, 6, and 10—DCBA; Subjects 3, 7, and 11—ABCD; Subjects 4, 8, and 12—BADC. Before each block of trials the subject was told what the nature of the signal was to be and was simply asked to press the key as quickly as possible upon hearing the signal. A verbal ready signal was given over an intercom approximately 2 seconds before each stimulus signal and the duration of the occurrence or absence of the tone was 2 seconds. The interval between trials was approximately 15 seconds.

The design of the experiment was such that the scores of the 12 subjects could be averaged for each trial irrespective of the conditions under which they were tested, so as to cancel out the main effects due to individual differences and to amount and direction of change in stimulus intensity as the signal. The adoption of this procedure left not only practice as a main effect but also *shifts* in the nature of the signal. Between Trials 9–10, 18–19, and 27–28 there were shifts in *degree* of change in stimulus intensity (small to large or large to small), but between Trials 18–19 there occurred a shift in *direction* of change in stimulus intensity (onset to offset or offset to onset) as well as in degree. Mean performance for the ordinal series of trials is depicted in Figure 4 and it is there clear that whereas the shifts in degree of change simply merged with the general practice effect, the shift in direction as well as degree caused considerable negative transfer. Wilcoxon's rank test was applied to a comparison of the scores on Trials 18 and 19 and this showed that the negative transfer was significant ($z = 3.43$, $p < .001$). These results are interpreted as clearly supporting the Perkins-Logan theory as against a modified stimulus-dynamism postulate.

The findings of the first experiment and other studies were also confirmed in the data. Working with the total reaction time over the last five trials in the block of nine for each subject, in order to reduce the practice effects, rank tests were used to assess the main effects of weak versus strong stimuli and of onset versus offset as a signal. The former effect was significant ($z = 3.19$, $p < .01$) but the

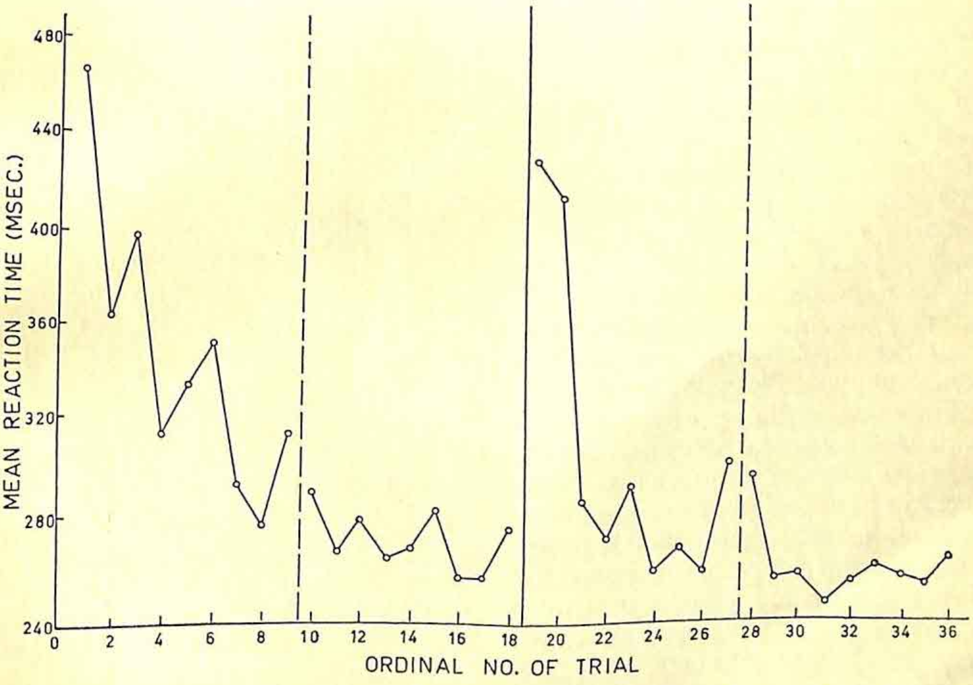


FIG. 4. Performance in Experiment II as a function of the ordinal number of the trial regardless of condition of testing. (Shifts in degree of change in stimulus intensity as the signal occurred between Trials 9–10, 18–19, and 27–28, but between Trials 18–19 there was also a shift in direction of change.)

latter effect was not significant ($z = 1.55$, $p = .12$). Thus, as Schwartz (1958) observed in the case of rats, strength of response varied with amount of change in stimulus intensity, regardless of the direction of the change.

The results of the second experiment nullified the attempt to modify Hull's treatment of the stimulus-intensity phenomenon and lent further weight to the Perkins-Logan theory. Apart from the direct disproof of the modified stimulus-dynamism principle, the differential conditioning theory should be preferred as the more parsimonious, for it avoids the assumption that a stimulus acts through change in intensity rather than its mere presence. This is not to deny rapid adaptation, so that a stimulus is most effective at the moment of onset, but it does imply that mere offset cannot be regarded as having stimulus properties. In terms of the Perkins-Logan theory, offset is effective because it inevitably means the simultaneous onset of some other physical energy as far as receptor processes are concerned. It should be noted that the terms "offset" and "onset" need not imply any direct intervention on the part of the experiment, as in the present experiments, but may be applied to changes in the energies affecting the receptors which are induced by the organism's own movements, as in instrumental conditioning or discrimination learning.

The use of such a precise and well-controlled response as key pressing by humans in this experiment has drawn attention to a point which may well demand further study if the differential conditioning theory is to find wide acceptance. The theory states that background or intertrial stimuli acquire inhibitory properties with respect to a given response when that response occurs in their presence without reinforcement. Although it was observed in Experiment I that GSRs commonly occurred between trials early in training, there were relatively few cases of key pressing to the intertrial stimulus in Experiment II. Some subjects who did respond before the signal of tone onset or offset apologized to the experimenter or acted as if they had behaved foolishly; the intertrial GSRs, on the other hand, would have passed undetected by the subject. Thus it may be necessary to assume that with such "voluntary" responses as key pressing it is enough for the response to occur implicitly without reinforcement for inhibition to accrue to the stimulus present at the time. This raises the further problem of the nature of the reinforcement for learning to give a "voluntary" response more and more quickly under instruction. No attempt will be made to enter into a discussion of such issues here, but it is clear from Figure 4 that learning did occur in a situation of this type and the immediate drop in response strength with a reversal of signaling and intertrial stimuli is in keeping with the Perkins-Logan theory as it is formulated at present.

Referring again to the comparison of the stimulus-intensity situation with that of discrimination learning, the results of Experiment II permit a parallel to be drawn. This is to the effect that in discrimination learning with white as the positive and black as the negative cue, negative transfer is obtained with a reversal to white as negative and black as positive after some training. In terms of Spence's (1937) analysis of the transposition phenomenon, furthermore, there should be some loss in performance with a shift from white to gray as the positive cue, with black negative throughout, or from black to gray as the positive cue with white constantly negative, because the difference between the positive and negative cues is less after the shifts; conversely, there should be gains in performance with shifts of the positive cue in the reverse direction, i.e., from gray to white with black negative or from gray to black with white negative, there being a greater difference between the positive and negative cues after the shifts. Comparable with these findings or expectations are the facts that the reversal of signaling and intertrial stimuli in Experiment II led to a marked drop in performance, and that level of performance depended on the degree of difference in intensity between these two stimuli. The latter feature of Experiment II did not show up in Figure 4, of course, because the design counterbalanced large and small intensity differences at successive shifts, but it was clearly evident when degree of change was studied as a main effect, as in Experiment I also.

STIMULUS INTENSITY AND DRIVE STRENGTH

The third and fourth studies of this series were conducted with a common aim, the determination of the general nature of the joint effect upon response strength of variations in stimulus intensity and drive strength. As noted at the outset, Hull *assumed* a multiplicative relationship between stimulus-intensity dynamism (V) and drive strength (D), but the Perkins-Logan theory should allow a *deduction* to be made in this respect. If the excitation accruing to the eliciting stimulus and the negative effects of generalized inhibition are set down in a conventional way, then the determination of the strength of the response in a variety of situations may be represented as

$$R = f(E) = f(H \times D - I)$$

where E is net excitatory potential, H is habit strength, D is drive strength, and I is inhibition (Spence, 1960). The application of this theorizing to the case of a response which is elicited by both weak and strong values of the stimulus, with H held constant, leads to the deduction that variations in the strength of the eliciting stimulus (i.e., in degree of inhibition) will act *additively* with variations in D. It is this deduction from the Perkins-Logan theory, opposed as it is to Hull's assumption, which was tested in the following two experiments. Because of its greater sensitivity to motivational effects and its "involuntary" nature, use was again made of the GSR. A factorial design was used to test for the additive or multiplicative relationship, it being assumed that the latter would be reflected in significant interaction between the variables.

Instead of using stimulus offset as the CS in a situation requiring new learning (as in Experiment I), capital was made of the fact that, due either to transfer from past experience or to the nature of the innate reflex, most subjects give a GSR to a tone without formal training, the response to a strong tone being greater than that to a weak tone. While many subjects also give GSRs between trials at first, the intertrial responses soon fall to a minimum frequency. The experiment therefore proceeded straight to an examination of the relation between this existing stimulus-intensity effect and drive strength. Variation in drive strength was achieved through the strength of shocks administered at random among the test presentations of the weak and strong tones, a technique used in a number of recent studies for this purpose (e.g., Spence, Haggard, & Ross, 1958). The repeated presentation of a tone as an eliciting stimulus for the GSR without reinforcement, however, produces the decrement in performance usually ascribed to either or both of the processes of adaptation and experimental extinction. In order to carry out a sufficient number of test trials (tone alone) it was necessary in one of the two experiments to interpolate training sessions among test sessions. The training sessions consisted of regular conditioning trials, with tone as CS and shock as UCS. The separate test series in this case each required an appropriate combination of tone strength and shock strength, but care was taken in the training series to ensure that each "boosting" session included all combinations of CS and UCS intensity in order that no artifacts would be introduced by the conditioning of the GSR to strength of tone as such or of the size of the response as such. By this means the assumption of constant habit strength for the GSR to the tone was justified.

The detail of the deduction to be tested under these conditions is as follows. Let E_{WW}, E_{WS}, E_{SW}, and E_{SS} represent the degree of net excitatory potential generated with the combinations of weak tone and weak shock, weak tone and strong shock, strong tone and weak shock, and strong tone and strong shock, respectively. Then

$$E_{WW} = H \times D- - \bar{I}+$$
$$E_{WS} = H \times D+ - \bar{I}+$$
$$E_{SW} = H \times D- - \bar{I}-$$

and

$$E_{SS} = H \times D+ - \bar{I}-$$

where H is constant, $D+$ and $D-$ are due to strong and weak shock, respectively, and $\bar{I}+$ and $\bar{I}-$ are the values of the generalized inhibition present with weak and strong tones, respectively, as eliciting stimuli. This state of affairs is represented in Figure 5.

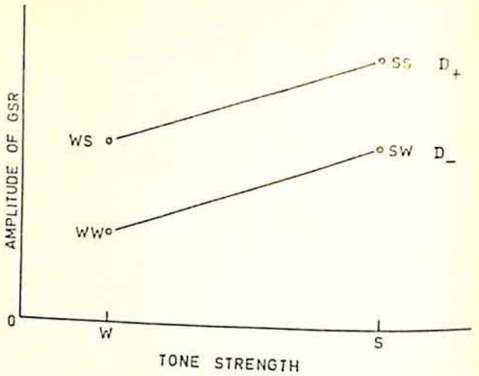


FIG. 5. The application of the Perkins-Logan theory to the covariation of stimulus strength (tone) and drive strength (shock). (The broken line indicates the gradient of inhibition, originating at the intertrial stimulus—*O*—and generalizing to the weak tone—$\bar{I}+$—and strong tone—$\bar{I}-$. The E values are for the four combinations of weak or strong tone—first subscript—with weak or strong shock—second subscript.)

The use of a factorial design allowed a test of an additive relationship in the absence of interaction between the variables. As set out in Figure 6, this test may be made by comparing the difference between the mean GSRs for strong and weak shock with a weak tone (WS-WW in Figure 6) with the difference between means for strong and weak shocks with a strong tone (SS-SW). The comparison may

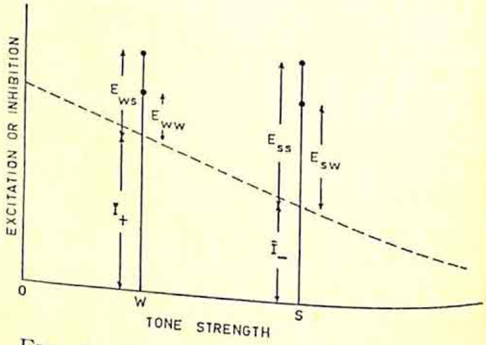


FIG. 6. Schematic representation of an additive relationship between stimulus strength (tone) and drive strength (D) generated through shock. (The letters indicate the four combinations of weak or strong tone—first letter—with weak or strong shock—second letter.)

also be made between the differences for strong and weak tone with a weak shock (SW-WW) and for strong and weak tone with a strong shock (SS-WS). The specific predictions from the Perkins-Logan theory are obtained thus:

$$\begin{aligned} E_{WS} - E_{WW} & \\ &= HD+ - \bar{I}+ - HD- + \bar{I}+ \\ &= H(D+ - D-) \end{aligned}$$

and

$$\begin{aligned} E_{SS} - E_{SW} & \\ &= HD+ - \bar{I}- - HD- + \bar{I}- \\ &= H(D+ - D-) \end{aligned}$$

when it is to be expected that the differences will be equal and that there will be an additive relationship such as is depicted in Figure 6. It also follows that both $E_{SW} - E_{WW}$ and $E_{SS} - E_{WS}$ are equal to $\bar{I}+ - \bar{I}-$.

EXPERIMENT III

The subjects were one man and one woman. The apparatus differed from that used in Experiment I in the following respects. The eliciting stimulus was a 2000-millisecond, 2000-cps sine-wave tone delivered to the subject through headphones at 60 db (weak tone), 75 db (medium tone), or 90 db (strong tone) SPL. The noxious stimulus used to induce drive was a 500-millisecond, 90-cps, spike-wave shock delivered through copper electrodes in the form of spring clamps $\frac{3}{4}$ inch wide which almost encircled the third and fifth fingers of the right hand near the fingertips, and it was set at 2.27 (weak shock), 3.20 (medium shock), or 4.80 (strong shock) milliamperes. After being placed in the apparatus the subject was merely told to keep still and warned of the first shock and the first tone. Both subjects underwent a number of test sessions at the rate of one each day at the same hour (11:00 A.M.–12:30 P.M.) and each session consisted of two shock trials, six tone and six shock trials in random order, and one tone trial. Shock strength and tone strength were held constant throughout any one test session but were varied from session to session in accordance with the basic aim of the experiment. The test sessions for the female subject involved the six combinations of weak (W), medium (M), and strong (S) tone with weak (W) and medium (M) shock, and the order of combinations over sessions was WW (weak tone and weak shock), SM, WM, SW, MW, and MM. This order did not seem to allow any biasing effect due to adaptation and if anything appeared to favor the occurrence of a direct multiplicative relationship. The test sessions for the male subject involved the four combinations of weak and strong tone with weak and strong shock. In this case it was possible to use the more desirable ABBA sequence and the combinations were used in the order SS, SW, WS, WW, WW, WS, SW, and SS. Twenty-four hours before each test session (following the previous test session) the subject was given a training series of nine conditioning trials (tone-shock with a 500-millisecond interstimulus interval) to combat adaptation and extinction effects, and each training series included all the combinations of tone strength and shock strength.

The data for each subject were treated separately and they are presented in Table 2. The results of an analysis of variance are set out in Table 3. In neither case did the tone-shock interaction, to be expected with a multiplicative relationship, prove statistically significant.

EXPERIMENT IV

The subjects were 20 volunteers (9 men and 11 women) from an introductory psychology course at the University of Sydney. With two exceptions the apparatus and stimuli were identical with those used in Experiment III. The copper shock electrodes were replaced with a type similar to those used to pick up the GSR, and only the 1.60-milliampere and 3.20-milliampere shocks were used, being des-

TABLE 2

MEAN GSR TO TONE (micromhos) IN EXPERIMENT III

Subject	Tone strength	Shock strength		
		Strong	Medium	Weak
Female	Strong	—	4.79	2.60
	Medium	—	2.74	1.26
	Weak	—	2.91	0.16
Male	Strong	1.30		0.27
	Weak	0.52	—	0.07

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF GSRs IN EXPERIMENT III

Subject	Source	*df*	*MS*	*F*
Female	Tone (A)	2	1797	3.24
	Shock (B)	1	4864	8.78**
	A × B	2	139	0.25
	Residual	36	554	
Male	Tone (A)	1	340	4.05*
	Shock (B)	1	757	9.01**
	A × B	1	109	1.30
	Residual	52	84	

* $p < .05$.
** $p < .01$.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF GSRs IN EXPERIMENT IV

Source	*df*	*MS*	*F*
Tone (A)	1	2220	42.69**
Shock (B)	1	152	2.92
Subjects (C)	9	243	4.67*
A × B	1	2	0.04
A × C	9	276	5.31*
B × C	[illegible]	233	4.52*
A × B × C	[illegible]	52	

* $p < .05$.
** $p < .01$.

ignated weak and strong, respectively, henceforth. The measures for each subject were taken in a series consisting of one medium tone trial, two shock trials, and, in a different random order for each subject, five weak tone, five strong tone, and eight shock trials. The subjects were matched in pairs on the basis of the amplitude of the GSR on the first tone trial (medium), one member of each pair then being tested with strong neighboring shock and the other with weak shock. Thus whereas shock strength was constant during the session for any one subject, both strong tone and weak tone trials were given to all subjects. Because each subject was tested in one session, with only 11 tone alone trials, adaptation and extinction effects were relatively slight and no training trials were used. Once again use was made of the facts that subjects enter this situation with a tendency to respond to the tone and that intertrial responses are soon inhibited. In view of the matching of subjects across shock strengths the data were treated as if the same subjects had undergone each combination of stimulus strengths and the results of an analysis of variance are shown in Table 4. Again the tone-shock interaction was not significant, as might be expected from an inspection of the data in Table 5.

The data of Experiments III and IV confirmed the deduction from the Perkins-Logan theory and were incompatible with the Hullian assumption of a multiplicative relationship between stimulus strength and drive strength. Another aspect of Walker's (1960) study, cited earlier, produced a similar finding. Half the subjects were trained with a weak UCS (air puff) and half with a strong UCS, and if this noxious UCS is thought of as a source of drive then her experiment also allowed a test of the relationship between stimulus intensity and drive strength. Walker found significant differences in acquisition due to both CS intensity and UCS intensity, but there was no interaction between these variables. This result cannot be taken as clear-cut, for the superior reinforcement of a strong UCS may have produced superior learning (Spence et al., 1958), as the strong CS may also have done. If either of these latter factors had produced differential habit strengths then a multiplicative relationship should have been obtained; the reason for this expectation emerges from a comparison with discrimination learning which follows shortly.

TABLE 5

MEAN GSR TO TONE (micromhos) IN EXPERIMENT IV

Tone strength	Shock strength	
	Strong	Weak
Strong	2.51	2.08
Weak	0.98	0.63

At least two other studies have produced evidence for Hull's formulation and some reconciliation must be made of these findings with the present data. First, with reference to an experiment on eyelid conditioning, Spence (1956) has reported a significant interaction between CS strength and anxiety level at asymptotic values of response strength. Second, Castaneda (1956) found the same type of relationship to hold for the latency and amplitude of a manual response, again with manifest anxiety as the drive variable. A common property of these two studies was the use of anxiety as a means of manipulating drive strength and the particular form of the data might be explained in terms of the "reactive" hypothesis of anxiety, i.e., with the idea that the difference between high and low anxious subjects from the motivational point of view is only apparent when they are presented with a "strong" eliciting stimulus or, in terms of the Perkins-Logan theory, with a stimulus which evokes little generalized inhibition. According to this view the multiplicative effect is obtained with anxiety as a drive variable because the effective D for the high anxious subjects is greater with a strong than with a weak eliciting stimulus. In the present experiments, on the other hand, the after effects of shock (D) are the same for strong and weak eliciting stimuli (tones), the reactive hypothesis does not apply, and the relationship under study takes the simpler additive form. It should also be noted that the experiment quoted by Spence was also complicated by the possibility that true learning was superior with the stronger CS.

It may be suggested that the rationale of Experiments III and IV was invalidated by the use of the GSR. In Experiment I it was found that effects predicted by the differential conditioning theory could be obtained with the GSR, so there can be no objection to its long latency as such, but a further difficulty may be seen in the fact that the GSR has often been used as a measure of emotion. Spence (1956) has assumed that an emotional response of an implicit nature (r_e) underlies the effects of aversive motivation (D) as generated by shocks, and the GSR might be taken as reflecting the strength of this emotional activity. If this argument be extended to allow that any stimulus, including a tone, arouses some emotional activity, then the use of the GSR in these experiments on the combined action of the tone and the aftereffects of shock might simply reflect the summation of two forms of emotional activity. There are at least two answers to this objection. In the first place, no theorist has supposed that a weak tone is motivating and it should follow that the granting of motivational properties to a strong tone only leads to the expectation of a multiplicative relationship between tone strength and shock strength. Second, it is likely that the latency of the implicit emotional or fear response is too long to affect even the relatively long-latency GSR. For the present purposes, therefore, it seems appropriate to treat the GSR as any other response within the framework of the differential conditioning theory, but to capitalize on its long latency in the manner indicated earlier.

Another point which must be considered is the methodological criticism recently levelled by Miller (1959) at the present type of study, which seeks the form of a relationship by simply taking representative values instead of covering the entire range of both variables. Miller's point is that the form of the relationship may vary with the location of the sampled values within the entire range, and this is true if one of the variables is the number of reinforced trials (N), which causes response strength to increase in the form of a negative growth function or some kind of nonlinear progression. In the present instance, however, both D and $\bar{I}$ are assumed to bear a linear relationship to shock strength and tone strength, respectively, and it should not matter what values are chosen for experimentation within limits set by the introduction of physical damage factors with extremely strong shocks and loud tones. Even if a negatively accelerated relationship were assumed to hold between $\bar{I}$ and tone strength, as is more likely, then the deduction of an additive effect made above should still be valid, for the differences in E values for weak and strong shock at various tone strengths are not affected by the nature of this relationship and should stay constant (Figure 5). Quite apart from this argument it is felt that the values used in the present experiments covered the range to a reasonable degree; as Table 2 shows, the weak tone-weak shock combination was so weak as to produce a barely measurable response in the male subject on most trials, the mean value being .07 micromhos, whereas the strong tone-medium shock combination produced large responses with a mean of 4.79 micromhos in the female subject.

A second methodological factor which might affect the validity of any attempt to determine the nature of the joint effect of two variables in a theoretical system is the unit of measurement to be used for the dependent variable. The deduction from the Perkins-Logan theory of an additive relationship between stimulus strength and drive strength (independent variables) as joint determiners of net excitatory potential (intervening variable) was tested through the amplitude of the GSR (dependent variable). The link between net excitatory potential (E) and amplitude of GSR thus becomes crucial if the additive effect deduced to apply to the former variable is to be reflected in the latter variable. It is fortunate, therefore, that evidence of a simple linear relationship between these two variables is available (Spence, 1956, pp. 113–116). The nature of the scales for quantifying H, D, and I need not enter into the question, for the expression in these terms distinguishing response strength at the two drive levels is identical for weak and strong eliciting stimuli. While amplitude of response may therefore be used with confidence as a measure in the test of the deduction from the differential conditioning theory, other measures of response strength may not be equally appropriate. The present state of behavior theory leads to the deduction that frequency of response is a normal ogival function of superthreshold excitatory potential, in which case the evidence of Spence (1956) and Walker (1960) with frequency of conditioned response (CR) as a measure of E may be suspect. Latency of response should theoretically bear the same relationship to E as amplitude as a measure, so that the data of Castaneda (1956) which were also cited above are less open to this criticism.

Further comparisons may be made with discrimination learning. In deriving the prediction of an additive relationship from the Perkins-Logan theory, use was made of the formulation

$$R = f(E) = f(H \times D - I).$$

It will be noted that D is assumed only to multiply H and not I. This is a long-standing assumption, used by both Hull and Spence, and it is supported by a wide variety of data. In the usual discrimination learning situation the learner is presented with two cue stimuli on each trial, with the response to one cue reinforced and to the other not reinforced. In these circumstances the positive cue should acquire habit strength (H) which generalizes to the negative cue ($\bar{H}$), while the negative cue acquires inhibitory potential (I) which generalizes to the positive cue ($\bar{I}$). If some uniform background is present throughout the experiment, and this background stimulus lies on the same dimension as the positive and negative stimuli, then it too will acquire inhibitory properties (I_B) which generalize to the positive stimulus ($\bar{I}_{B+}$) and to the negative stimulus ($\bar{I}_{B-}$). The application of the formula to give the net excitatory potential of the positive stimulus ($E+$) and the negative stimulus ($E-$) then gives the following expressions:

$$E+ = H \times D - (\bar{I} + \bar{I}_{B+})$$

and

$$E- = \bar{H} \times D - (I + \bar{I}_{B-})$$

when it follows that, if the two E values under high drive ($D+$) be represented as $E++$ and $E-+$, and under low drive ($D-$) as $E+-$ and $E--$,

$$E++ - E-+ = D+(H - \bar{H}) - k$$

and

$$E+- - E-- = D-(H - \bar{H}) - k$$

where k is a constant representing the total of the four sources of inhibition. Now $D+$ is greater than $D-$, so it has been deduced that discrimination learning under the stipulated conditions should be superior with high drive. The other derivation analogous with that made for the stimulus intensity situation requires a comparison of the E values for the positive stimulus with high and low drive ($E++ - E+-$) and the negative stimulus with high and low drive ($E-+ - E--$), as follows:

$$E++ - E+- = H(D+ - D-)$$

and

$$E-+ - E-- = \bar{H}(D+ - D-)$$

and since H is greater than $\bar{H}$, because the latter has generalized from the former, it is to be expected that there will be interaction between the degree of separation of the stimulus cues on the one hand and drive level during training, on the other hand. Reference back to the deductions about the joint effects of stimulus intensity and drive strength which were tested in Experiments III and IV will show that the assumption of constant habit strength is there crucial, for any difference in this respect between strong and weak eliciting stimuli will lead to a multiplicative rather than an additive relationship with drive strength. Of equal importance is the limitation of drive effects to habit strength and the supposition that drive does not multiply inhibition; modification of the basic formula to read

$$R = f(E) = f(D \times H - D \times I)$$

would also lead to the prediction of a multiplicative effect.

In a recent review of drive effects in spatial discrimination learning, Brown (1961) has reached a conclusion which tallies exactly with the deductions made above, i.e., that when the habit strengths of the response tendencies for the positive and negative cues are held equal, drive differences do not affect percentage choice of the correct response; if a correction method of learning is used, or if the two responses are reinforced with differential frequencies, then drive effects become apparent. Brown notes, however, that there are too few relevant data to allow any such firm conclusion in the case of nonspatial discrimination learning. If anything in the latter case, performance improves with increased motivation, and this may be due to the greater difficulty in keeping habit strengths equal in the nonspatial situation with forced trials. It should also be noted that many of the studies cited by Brown involve instrumental approach learning, and that under these conditions the mere occurrence of the response, with or without reward, is assumed to produce an increment of habit strength; for this reason, forced runs to the nonrewarded side of a T maze, for example, are thought to produce as much learning as a free run to the rewarded side, and the increasing preference for the rewarded side is explained in terms of incentive motivation (Spence, 1956). It is therefore as well to emphasize the earlier caution that the analogy between the mechanism of the Perkins-Logan theory and discrimination learning is inexact, and that stimulus-intensity effects are more safely treated as due to differential classical conditioning.

Before concluding this discussion of the interrelated effects of stimulus intensity and drive strength, attention must be drawn to some recent comments by the author on drive effects in conflict (Champion, 1961). In order to attempt a resolution of an apparent inconsistency in the S-R treatment of conflict, it was assumed that changes in the strength of separate drives, such as hunger and fear, not only produce a common result by contributing to overall drive level, according to the generalized drive hypothesis, but also lead to distinguishable effects of a selective nature through the dynamism of the drive stimuli. Thus it was supposed that an increase in the hunger of a rat in approach-avoidance conflict affects the generalized habit strength responsible for the gradients of both approach and avoidance (generalized drive) and also acts selectively on the approach gradient through the strength of the hunger-drive stimulus. Although the evidence presented in the present account of stimulus-intensity effects would seem to force one to discard the *assumption* of a stimulus-dynamism principle, the *facts* of stimulus-intensity changes remain unchallenged and it is still possible to hypothesize that a stronger drive stimulus elicits a stronger response, other things being equal. As for the *explanation* of such a process, the substitution of the dimension of drive strength for that of the intensity of the eliciting external stimulus may pose further problems. It would have to be assumed that the response under study is not reinforced when the organism is satiated, so that satiation stimuli take on inhibitory properties. This latter inference is similar to some recent proposals by Estes (1958). More generally, however, the present line of argument preserves the purely motivational or energizing property of drives, while allowing them stimulus properties as well, whereas Estes appears to be moving towards a purely associative explanation of drive effects.

An experimental test of these alternatives might be conducted along the following lines. Suppose that a group of rats was trained to perform some act such as lever pressing at 4 hours food deprivation; the use of some reinforcement other than food would make this learning technically easier to obtain. Suppose also that the same animals were regularly placed in the apparatus at 12 hours deprivation but that the response of lever pressing was never reinforced at this level. Considering only the role of the drive stimuli and assuming that the effect of drive is purely associative, the application of the Perkins-Logan theory to this form of differential training is set out in Figure 7. This figure differs from preceding representations in that habit strength rather than gross excitatory potential is plotted on the ordinate, along with inhibition. The lever pressing response is reinforced at 4 hours deprivation, and 6 units of habit strength are arbitrarily assigned to the drive stimuli at that level ($H = 6$); because of drive stimulus generalization, 5 and 4 units of generalized habit strength ($\bar{H}$) are assigned to the drive stimuli at 8 and 12 hours deprivation, respectively. The response is also allowed to occur at 12 hours deprivation, but it is never reinforced at that level, so that enough inhibition (I) accumulates to cancel the generalized habit strength, viz., 4 units; through drive stimulus generalization again, 3 and 2 units of generalized inhibition ($\bar{I}$) are assigned to the drive stimuli at 8 and 4 hours deprivation, respectively. Two methods of calculating net excitatory potential are then compared in Table 6. If a purely associative theory of drive is used (Figure 7), then net excitatory potential (E) should decrease with deprivation time, as shown in Table 6 when $E = H - I$. If, on the other hand, deprivation brings generalized energizing effects (D) as well as drive stimuli (S_D), and if $E = H \times D - I$, then more inhibition will have to be generated to stop the response occurring at 12 hours deprivation, and net excitatory potential should reach a maximim at some period of deprivation between 4 and 12 hours (Table 6).

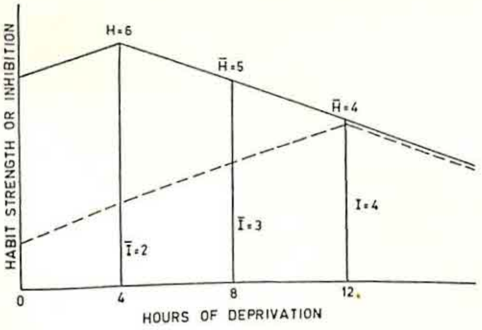


FIG. 7. The application of the Perkins-Logan theory to a hypothetical situation in which a given response is reinforced at 4 hours deprivation of food but not reinforced at 12 hours deprivation. (It is assumed in this instance that the effects of drive are purely associative.)

TABLE 6

ALTERNATIVE DERIVATIONS OF NET EXCITATORY POTENTIAL FROM VALUES ARBITRARILY ASSIGNED IN FIGURE 7

	Hours of deprivation		
	4	8	12
Habit strength (H or $\bar{H}$)	6	5	4
Drive strength (D)	2	4	6
$H \times D$	12	20	24
If $E = H - I$			
I equals	2	3	4
E equals	4	2	0
If $E = H \times D - I$			
I equals	12	18	24
E equals	0	2	0

CONCLUSIONS

Taken collectively, the results of the present experiments confirm the Perkins-Logan theory of stimulus-intensity effects in terms of differential conditioning and are incompatible both with Hull's original formulation and the modification proposed by the author. Thus the feature of the stimulus situation which seems to be directly responsible for the phenomena under study is the relative disposition of the eliciting stimulus and the background or intertrial stimuli on some common dimension of generalization. The Perkins-Logan theory is also meaningful in the sense that it permits the generation of a number of predictions which remain to be tested. The denial of Hull's stimulus-intensity dynamism postulate and the verification of the differential conditioning theory, however, does not jeopardize the general principles of the type of behavior theory which he and his co-workers established. On the contrary, the differential conditioning theory follows directly from some general features of Hull's behavior system, based as it is on the accepted principles of excitation following reinforcement, inhibition following nonreinforcement, stimulus generalization, and the algebraic summation of excitation and inhibition. What Perkins and Logan have done is imply that an *assumption* of stimulus-intensity dynamism is redundant within the general framework of Hull's behavior theory, and the empirical data are in keeping with this implication. The third and fourth experiments described in this paper went one step further, however, in showing Hull's assumption of a multiplicative relationship between stimulus intensity and drive strength to be invalid as well as redundant.

SUMMARY

Following an outline of the Hullian and Perkins-Logan theories of stimulus-intensity effects, and a distinction between the role of these effects in learning and performance, four experiments were described which tested the relative merits of the two theories in situations limited to effects on performance. In the first experiment it was found that, with stimulus offset as a CS, response strength late in training was directly related to the strength of the offset stimulus. Because this finding could be readily accounted for by a modification of Hull's stimulus-intensity dynamism postulate couched in terms of degree of stimulus change as the effective variable, regardless of the direction of the change, the second experiment was designed to detect any negative transfer occurring with a shift from onset to offset or vice versa in the one group of subjects. Negative transfer did occur and this outcome was taken to be incompatible with the modification of Hull's postulate and as further evidence for the Perkins-Logan theory in terms of differential conditioning. In the third and fourth experiments tests were made of a deduction from the latter theory to the effect that there should be an additive relationship between stimulus strength and drive strength in their joint determination of response strength, contrary to Hull's assumption of a multiplicative relationship. The two studies provided evidence for an additive relationship. It was concluded from the series of experiments as a whole, that the Perkins-Logan theory is to be preferred to Hull's treatment of stimulus-intensity effects. Some other implications of the former theory were also considered, mainly through comparisons with discrimination learning.

REFERENCES

BROWN, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

CASTANEDA, A. Reaction time and response amplitude as a function of anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1956, **53**, 225–228.

CHAMPION, R. A. Motivational effects in approach-avoidance conflict. *Psychol. Rev.*, 1961, **68**, 354–358.

ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958. Pp. 35–68.

GRANIT, R. *Sensory mechanisms of the retina.* New York: Oxford Univer. Press, 1947.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, 38, 690–696.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, 39, 35–40.

GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, 40, 702–708.

HANSCHE, W. J., & GRANT, D. A. Onset versus termination of a stimulus as the CS in eyelid conditioning. *J. exp. Psychol.*, 1960, 59, 19–26.

HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, 56, 67–77.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

LOGAN, F. A. A note on stimulus intensity dynamism (*V*). *Psychol. Rev.*, 1954, 61, 77–80.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

NYGAARD, J. E. Cue and contextual stimulus intensity in discrimination learning. *J. exp. Psychol.*, 1958, 55, 195–199.

PERKINS, C. C., JR. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, 46, 225–231.

RABEN, M. W. The white rat's discrimination of differences in intensity of illumination measured by a running response. *J. comp. physiol Psychol.*, 1949, 42, 254–275.

SCHWARTZ, M. Conditioned-stimulus variables in avoidance learning. *J. exp. Psychol.*, 1958, 55, 347–351.

SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol Rev.*, 1937, 44, 430–444.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs: Prentice-Hall, 1960.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, 55, 404–411.

WALKER, E. G. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1960, 59, 303–311.

(Received March 3, 1961)

Psychological Review
1962, Vol. 69, No. 5, 450–466

CONDITIONING THROUGH VICARIOUS INSTIGATION

SEYMOUR M. BERGER[1]
Indiana University

Personality theorists and social psychologists generally recognize that the emotional responses of one person (performer) may elicit emotional responses from another (observer). When these emotional responses are similar, the relationship between the performer and observer is described as *empathetic,* or one of *identification.* If the relationship is one of *envy,* the performer's pleasure elicits an unpleasant emotional reaction from the observer. In a *sadistic* (malicious joy) relationship, the performer's displeasure elicits a joyful response from the observer. In all of these relationships, the observer's emotional response is dependent upon the emotional response of the performer. As social phenomena, empathy, envy, and sadism require that the observer experience the emotion of the performer *vicariously,* in the sense of perceiving it, as a prelude to the observer's own emotional response.

The purpose of this paper is to provide a conceptual framework within which to study empathy, envy, and sadism as determinants of interpersonal behavior. For example, empathy may account for certain forms of altruism, envy for the loss of friendship, and sadism for malicious mischief. Before proceeding to these problems, however, it is necessary to identify a basic mechanism in the arousal of the observer's emotional response.

VICARIOUS INSTIGATION

If an observer responds emotionally to a performer's unconditioned emotional response (UER) then vicarious instigation has occurred.[2] The performer's UER requires more precise definition. It is meant to refer to the performer's emotional state, following the presentation of an unconditioned stimulus (UCS), *as perceived by the observer.* The actual state of the performer is not directly observable. The observer must infer the performer's UER from observable events in the environment. Typically, the observer will rely upon the UCS and the performer's overt unconditioned response (UCR) as cues to the performer's UER. Consequently, the performer's UCS and UCR are important determining factors in vicarious instigation.

The performer's UER need not actually occur, the observer may infer it (incorrectly) from the situational context. It is only necessary that the observer perceive that the performer's UER has occurred, and respond emotionally, for vicarious instigation. A mother, for example, may react with fear and anguish when she sees her child fall, only to find that the child is unhurt and undisturbed. In this example, the mother's emotional response (ER) was dependent upon her perception of what had happened to her child (the UCS) rather than the child's reaction (UCR). She misperceived the child's emotional state in this case, yet she was vicariously instigated; she inferred the child's UER from the

---

[1] The author wishes to express his indebtedness to at least two colleagues, Marvin Levine and Leon H. Levy, for their encouragement, many helpful suggestions, and critical reading of the manuscript.

[2] Throughout this paper, the term *unconditioned* refers to dependable S-R relationships which serve as a basis for conditioning. This usage is in agreement with the definition employed by many writers in the area of conditioning (e.g., Hilgard & Marquis, 1961, pp. 46–47). In this procedural sense, the term unconditioned may refer to either learned or unlearned S-R relationships.

UCS only. Had she been in another room and heard the child cry (UCR), the mother might have inferred that her child was in pain (UER) without knowledge of the precipitating event (UCS), and yet would have been vicariously instigated. While the performer's UCS and UCR are important cues for vicarious instigation, both are not always necessary for vicarious instigation.

Pseudovicarious Instigation

The simple contiguity between the performer's UCS and UCR and the observer's ER is not sufficient evidence for vicarious instigation, however. Since the observer's ER follows the performer's UCS and UCR, the observer may appear to be responding to the performer's UER when in actuality he is responding to some other stimulus component in the situation. As indicated above, vicarious instigation is dependent upon the performer's UCR or the UCS. In some situations, however, the observer may respond to the UCS only, so that the performer's UER is superfluous. An observer may respond with fear to the firing of a gun, regardless of whether the target is a bull's-eye or another human being. The observer's fear is not vicariously instigated in this instance—although it may appear to be if the target happens to be another human being.

Another possible source of pseudovicarious instigation is the performer's UCR. The performer's UCR may be a UCS which elicits an observer ER; in this case, the observer responds to the performer's UCR only, so that the performer's UCS and UER are superfluous. A sudden loud scream by a performer, for example, may elicit a fear response from the observer independently of his response to the performer's UCS or UER. Recent studies by Church (1959) and Murphy, Miller, and Mirsky (1955) have demonstrated how a rat's squeal and a monkey's presence, respectively, may serve as a conditioned fear stimulus.

The performer's UER may serve as a discriminative stimulus which indicates to the observer that he will receive the UCS that is presented to the performer. This is another source of pseudovicarious instigation. One child may fear that he will be spanked when he sees that his brother is being spanked for no apparent reason. It is not his brother's cries in response to the spanking that elicit the child's fear, rather it is the anticipation of being spanked himself.

The following study by Barnett and Benedetti (1960) illustrates a form of pseudovicarious instigation which may occur in conditioning. In this case, an extraneous stimulus in the situation served as a UCS:

> After adapting the GSR of an observer to a buzzer, the authors had him watch a performer undergoing GSR conditioning, with shock as the UCS and the buzzer as the conditioned stimulus (CS). Even though most observers were not aware that the performer was being shocked, they responded with GSRs to the buzzer on test trials. Subsequent experimentation concerning the various stimulus components of the situation led the authors to conclude that what appeared to be "vicarious conditioning" (their term) involved responses to visible needle deflections of the galvanometer, even in the absence of the performer; the GSRs, in this case, reflecting changes in activation level following the occurrence of a stimulus change, i.e., needle movements.

All of these situations are not instances of vicarious instigation, although they would appear to be in casual observation. Obviously, therefore, anecdotal evidence for vicarious instigation is of dubious value, and an experimental analysis is required to eliminate the possibility of pseudovicarious instigation.

Concordant and Discordant Forms[3]

The definition of vicarious instigation requires that the observer respond emotionally to the performer's UER, but there is no restriction placed upon the nature of these emotional responses. As indicated previously, there may be a certain similarity or dissimilarity between the emotional responses of the performer and observer. Concordant vicarious instigation takes place when performer and observer responses are similar[4]; discordant vicarious instigation occurs when performer and observer responses are dissimilar. An examination of concordant and discordant forms of vicarious instigation provides a rudimentary classification system for empathy, envy, and sadism.

If discussion is limited to the elicitation of emotional responses which may be classified as either pleasant (+) or unpleasant (−), it is possible to specify the necessary conditions for empathy, envy, and sadism. The four possible combinations of pleasant and unpleasant emotional responses for performer and observer are presented in Table 1. It should be noted that throughout this discussion, the classification of the performer's UER is in terms of the observer's perception of it.

TABLE 1

COMBINATIONS OF EMOTIONAL RESPONSES FOR PERFORMER AND OBSERVER

Case	Performer's UER	Observer's ER
I	+	+
II	+	−
III	−	+
IV	−	−

Empathy may be defined by Cases I and IV; shocking the performer (−) may be painful (−) for the observer, for example. *Envy* may be defined by Case II; rewarding the performer (+) may instigate a feeling of depression in the observer (−). *Sadism* may be defined by Case III; shocking the performer (−) may elicit joy in the observer (+).

Although these combinations of performer and observer responses seem to identify the *necessary* conditions for characterizing empathy, envy, and sadism, the present scheme does not include the *sufficient* conditions for complete characterization.

Nevertheless, this initial effort suggests a useful direction to pursue in order to bring a wider variety of emotional responses within the scope of conditioning theory.

Conditioning

As a motivational concept, vicarious instigation has relevance to a variety of social behaviors acquired through classical and instrumental conditioning. In the discussion which follows, an attempt is made to identify two conditioning models in which social learning is a function of vicarious instigation.

---

[3] Although not specifically concerned with the problem of conditioning, Heider (1958, Ch. 11) has presented an analysis which describes one person's reactions "to the lot of the other." There are some parallels between the analysis of conditioning through vicarious instigation and Heider's presentation. The present analysis attempts to deal with similar problems, but in terms of conditioning models.

[4] Some authors (e.g., Asch, 1952, p. 171; Heider, 1958, pp. 279–282) distinguish between two forms of concordant emotional response. The observer may react concordantly with the performer's response to the unconditioned stimulus (UCS) and in a manner that is concordant with the way that the observer would react if he had been directly instigated; the observer may react concordantly with the performer's response to the UCS, but in a manner that is discordant with the way the observer would react if he had been directly instigated. In the former instance, the observer is reacting emphathetically, while in the latter instance the observer is reacting sympathetically. This distinction, while useful for some purposes, will be disregarded in this paper because it does not advance the present analysis.

CLASSICAL CONDITIONING PARADIGM: THE PERFORMER'S UER AS A UCS

In classical conditioning, in order for a response to become conditioned to a neutral stimulus (CS), the response must be elicited by the UCS. Any stimulus may serve as the UCS if it will elicit a given response with some degree of regularity. In vicarious instigation, the performer's UER serves as a UCS for the observer's ER.

The distinction between vicarious instigation and conditioning may be illustrated as follows: Haner and Whitney (1960) had an observer watch a performer who supposedly was being shocked each time a light was presented. The authors found that this situation elicited GSRs from the observers, the GSR size being positively related to the general anxiety level of the observer. But no test was made to determine whether the observer's GSRs became conditioned to the light alone. While Haner and Whitney call this phenomenon "empathic conditioning," the term vicarious instigation seems more appropriate.

In the classical conditioning model, vicarious instigation must occur in contiguity with a CS. Conditioning occurs when the CS comes to elicit the observer's ER in the absence of the performer's UER. While this model may be applied to a variety of social learning situations, an example of concordant vicarious instigation in the formation of attitudes may be found in this excerpt from Gordon W. Allport (1935):

> Even before he has an adequate background of appropriate experience, a child may form many intense and lasting attitudes toward races and professions, toward religion and marriage, toward foreigners and servants, and toward morality and sin. A parent's tone of voice in disapproving of ragamuffins who live along the railroad track is enough to produce an uncritical attitude in *the child who has no basis in his experience for the rational adoption of the parent's point of view* [italics added]. It frequently happens that *subsequent* experience is fitted into the attitude thus uncritically adopted, not . . . made the basis for the attitude. In such cases every contact is prejudged, contradictory evidence is not admitted, and the attitude which is *borrowed secondhand* [italics added] is triumphant (p. 811).

In terms of the model, the "ragamuffins" are the CS, the parent's UER is inferred from the disapproving tone of voice, and the child's uncritical negative attitude is the observer's ER. Once the emotional component of the attitude is conditioned ("borrowed secondhand"), the cognitive elements are added by subsequent experience.

Although attitudes are formed in a variety of ways, the present model provides a conceptual and experimental basis for studying the transmission of ready-made attitudes through observational learning in a classical conditioning framework. One can only speculate concerning the variety of social experiences that may be transmitted in this way.

INSTRUMENTAL CONDITIONING PARADIGM: VICARIOUS INSTIGATION AS A REINFORCEMENT

Vicarious instigation may serve as either a positive or negative reinforcement for observer responses in instrumental conditioning situations. Each type of reinforcement will be considered separately.

As a positive reinforcement, vicarious instigation strengthens the observer responses it follows. It is assumed that a positive ER is instigated in the observer. Therefore, pleasant empathetic feelings or malicious joy would be the basis for the positive reinforcement. Altruism, for example, is assumed to be based upon concordant vicarious instigation where both performer and observer emotional responses are positive. To illustrate, I may become more charitable when my gift is received with obvious pleasure by the recipient. If perceiving the pleasure of the recipient is vicariously instigating, it may reinforce my charitable behavior. Of course, sources of pseudovicarious instigation would have to be controlled, since charitable behavior may be reinforced by approval I receive from the recipient or from my family and friends.

"Tattletale" behavior is an example of how a sadistic relationship may affect conditioning. One child may learn that he can have his rival punished by disclosing the rival's misdeeds to his mother. Presumably, the enjoyment elicited by this rival's punishment reinforces the tattletale behavior. Again, experimental analysis is necessary to eliminate pseudoeffects; e.g., the child may be seeking the approval of the mother and a raise in his own status.

Vicarious instigation may function as a negative reinforcement if responses are strengthened when followed by the removal or reduction of vicarious instigation. If it is assumed that emotional states have motivating properties then, following Miller's (1948) paradigm, vicariously instigated emotions may function to initiate learning; furthermore, the observer will acquire responses which reduce the emotional drive stimulus.

The performer's UER may instigate the observer's ER (assumed to be negative). The observer's ER is the drive stimulus which instigates other responses. If one of these other responses is followed by the removal of the performer's UER, vicarious instigation is terminated, and so is the observer's ER. The response which was followed by the termination of vicarious instigation will be strengthened.

This model seems applicable to situations when the vicariously instigated emotion is negative. Consequently, both unpleasant empathetic feelings and feelings of envy can function as drive stimuli in this model. Consider, for example, the starving beggar who comes to the door for a handout. The housewife may feel grief at the sight of the emaciated beggar and upon hearing his plea. She may be able to reduce her grief by shutting the door to remove the pitiful figure from her view. In this case, the beggar's state (performer's UER) aroused the housewife's grief (observer's ER) which was reduced by shutting the door. Presumably, the response of shutting the door was effective in reducing the feeling of grief and was reinforced. There are, of course, other potential sources of reinforcement in this type of situation. For example, the housewife may feel threatened by the presence of a stranger at her door, and may shut the door to insure her own safety. Clearly, sources of pseudoeffects must be ruled out before it can be concluded that the behavior was a function of vicarious instigation.

Envy may also function as a drive stimulus in this model of conditioning. One man may be envious when his rival is promoted, and attempt to reduce his feelings of mortification by subsequently avoiding all encounters with his rival. Again, it should be noted that an analysis for pseudovicarious instigation must be made (e.g., the man may be responding to his loss of the promotion rather than his rival's good fortune) before concluding that vicarious instigation is the basis for the behavior.

A number of writers (e.g., Berger, 1961; Campbell, 1961; Hill, 1960; Sechrest, 1961) have used the term *vicarious reinforcement* to describe learning phenomena. The present model identifies vicarious instigation as a possible basis for vicarious reinforcement.

AN EXPERIMENTAL DEMONSTRATION OF VICARIOUS INSTIGATION IN CLASSICAL CONDITIONING[5]

The foregoing analysis has employed anecdotal evidence and hypothetical examples to illustrate and support the conceptual framework. In light of the discussion of pseudovicarious instigation, an experimental analysis is imperative, yet lacking. Since vicarious instigation is crucial to the framework, demonstrations seemed in order. As a starting point, the author selected the classical conditioning model as a basis for demonstrating vicarious instigation in learning.

Although developed independently, the general setting for the experiments to be reported is similar to the situations used by Haner and Whitney (1960) and Barnett and Benedetti (1960). In this situation, an observer watches a performer (actually a confederate of the experimenter, who supposedly is being shocked every time a light dims. The dimming of the light is preceded by a buzzer. The GSRs of the observer are recorded as a measure of his ER. Everytime the light dims and the performer is supposedly shocked, he jerks his arm from the inductorium.[6] The buzzer was the CS. The performer's UER was inferred from the dimming of the light and electric shock (performer's UCS) and the performer's arm movement (performer's UCR). Presumably, if the observer responded with a GSR on conditioning trials, vicarious instigation occurred. It was necessary, of course, to introduce pseudovicarious instigation controls.

The previous discussion of pseudovicarious instigation suggested three general sources of pseudovicarious instigation: The performer's response, per se; the performer's UCS, per se; and the observer's anticipation that the UCS will be presented to him.

EXPERIMENT I

In an initial experiment, an attempt was made to obtain information concerning the effectiveness of the experimental situation for vicarious instigation, and to determine if the instigation of observer GSRs by apparent shock to the performer was due to the performer's response, per se. The experiment contained two conditions:

1. Shock-Movement (S-M)—the observer was told that the performer would be shocked and the performer reacted with an arm movement when apparently being shocked.
2. No Shock-Movement (NS-M)—the observer was told that the performer would make a voluntary arm movement at a given signal (light dimming), and that the performer was not being shocked.

If no differences were found between these conditions, then it would appear that observers were responding solely to the performer's movement, and that the performer's UER was not a determining factor. If differences were found, where S-M observers were more responsive, then some support would be provided for the phenomenon of vicarious instigation.

Fifteen subjects were assigned at random to each condition. The performer (a male confederate of the experimenter who appeared to be just another subject recruited for the experiment) was the same for all subjects. In order to avoid any systematic bias in the performer's behavior, the instructions regarding the different conditions were read to the observer before the performer appeared for the experiment. For observers in the S-M condition, the experimenter explained that another subject would be shocked in the last part of the experiment, and that he would be a few minutes late because the experimenter's assistant was testing the other subject to see what level of shock he could take; for observers in the NS-M condition, the performer's initial absence was explained on the basis that the experimenter's assistant was giving the other subject practice in making the arm movement. Instructions, which were common to both conditions and did not reveal the specific nature of the condition, were read aloud when performer and observer were present together.

---

[5] The author wishes to express his appreciation to Indiana University for an equipment grant in support of this research, and to Neil Schneiderman, Susan Kroener, and Patricia Stewart who were the performers and scorers in these studies. The experimenter for Experiment I was the author, Patricia Stewart was the experimenter in Experiment II, Susan Kroener was the experimenter in Experiment III.

[6] The performer jerked his arm rapidly, as a twitch. The performer tried not to make other movements, or change facial expression.

The experiment was described to the observer as an attempt to determine whether different kinds of stimulation (buzzer and light diming) produced different types of skin resistance changes. S-M observers were told that the performer would be shocked every time the light dimmed in order to determine the effect of shock on the performer's GSR. The observer was reassured that he was a control subject and would not be shocked at any time. NS-M observers were told that the performer would try to jerk his hand off the apparatus in front of him every time the light dimmed to see if the movement affected his GSR. The observer was told that he was a control subject and that no one would be shocked.

Although the experimenter appeared to be taking readings from both the performer and observer, only the observer's GSRs were recorded. After an initial adaptation period, in which the buzzer was presented for 1.0 second followed immediately by dimming of the light for 1.0 second, the performer's hand was strapped to the inductorium in front of him. It was only during this latter phase of the experiment that the performer made arm movements.

There were 13 trials during this phase of the experiment. Ten trials were used for conditioning, with 3 test trials interspersed in the same position for all subjects. The intertrial interval was 15 seconds.

The method of moving averages with blocks of three trials was used for all curves presented in this paper, using the number of observers giving a GSR on given trials as the dependent response measure. For purposes of statistical analysis, the total number of GSRs for each observer, for a given set of trials, was used as the dependent response measure in all experiments.

Three criteria were used to identify a GSR response:

1. The movement of the recording pen had to be in the direction of lowered resistance.
2. The movement had to be at least as large as twice the thickness of the recording pen line.
3. The movement had to occur within 4 seconds following the termination of the buzzer.

The records were scored independently by the experimenter and the confederate, although both had some knowledge as to which conditions given observers had been assigned. The scorers agreed on 94% of the trials concerning whether or not a GSR had occurred; disagreements were resolved by re-examination and discussion by the scorers.

The criterion for adaptation during the first phase of the experiment was two successive no response trials; all observers were given a minimum of four adaptation trials, however. The adaptation curves over the four trials are presented in Figure 1. While it is apparent from Figure 1 that subjects in the S-M condition were less responsive than subjects in the NS-M condition, a comparison of the total number of individual responses over the four adaptation trials for each condition by the Mann-Whitney *U* test revealed no reliable differences ($U = 81.5$).[7]

The relative effectiveness of the two conditions to instigate observer GSRs may be examined in the "conditioning" section of Figure 1. The total scores for individual observers in the two conditions (on trials in which the confederate responded) were compared by the Mann-Whitney *U* test. S-M observers appear to be more responsive

---

[7] Two-tailed tests are used throughout this paper. The Mann-Whitney test was employed in Experiment I because the comparison groups differed significantly in variance (uncorrectable by usual transformation techniques) and, at the same time, appeared to have differently shaped distributions. The data for Experiments II and III were more amenable to analysis of variance according to the criteria suggested by Boneau (1960).

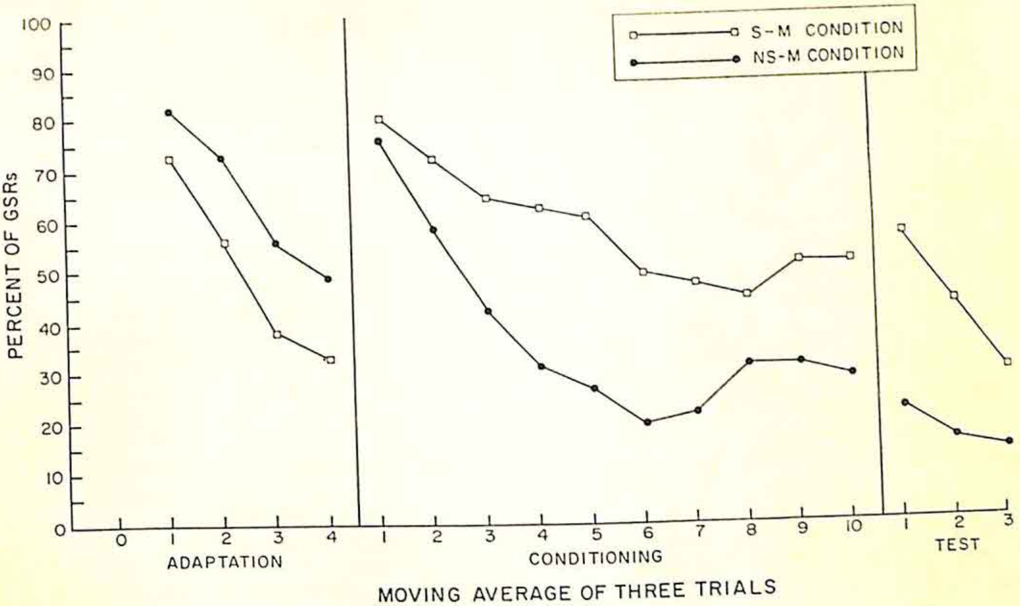


FIG. 1. Percentage of observers giving GSRs during adaptation, conditioning, and test trials (Experiment I).

than NS-M observers, but the difference is not reliable ($U = 70.5, p < .10$).

The total number of GSRs over the three test trials was obtained for each observer and the Mann-Whitney U test was used to compare the differences between the conditions. The "test trial" section of Figure 1 shows that S-M observers gave more responses on test trials than NS-M observers. The total score differences were reliable at less than the .05 level ($U = 61.5$).

A post-experiment questionnaire revealed that most observers gave descriptions of the performer's situation which were in agreement with the experimental instructions. Two NS-M observers reported that the performer might be receiving a shock. One observer in each condition reported some expectation of receiving shock.

These results lend support to the notion that conditioning can occur through vicarious instigation. While the arousal effects of the arm movement are apparent from the data, there is some tendency for GSR instigation to occur more frequently when the observer is told that the performer is being shocked. Furthermore, the frequency of conditioned responses (CRs) is reliably greater in the S-M condition. Conditioning was not solely a function of the performer's response, but was dependent upon the instructions concerning shock.

EXPERIMENT II

The encouraging results of Experiment I testified to the utility of the experimental setting in studying vicarious instigation. Yet, Experiment I contained only two pseudovicarious instigation controls; i.e., the NS-M condition, for performer response, and the instructions to the observer that he would not be shocked. A third control is necessary to determine whether the performer's UCS per se, will account for the differences obtained. It is conceivable that S-M observer's were affected solely by the instructions concerning shock, and that the performer's response was not a contributing factor; or, that the apparent presence of electricity in the inductorium had arousal effects which were independent of the performer *being* shocked. Since the differences between the movement conditions for instigations was not reliable in Experiment I, a complete 2 × 2 factorial design involving shock and movement was employed in this experiment. This design provided for the various pseudovicarious instigation controls. The four conditions of this experiment are identified as follows:

1. S-M—Shock to performer, and performer arm movement.
2. S-NM—Shock to performer, but no performer arm movement.
3. NS-M—No shock to performer, but performer arm movement.
4. NS-NM—No shock to performer, and no performer arm movement.

The S-M and NS-M conditions were the same conditions used in Experiment I. The S-NM condition is a "UCS control" for the instructions concerning shock and an electrified inductorium; the performer made no response. The NS-NM condition was another UCS control; the performer's hand was not placed on the inductorium during the conditioning phase of the experiment. She was told that she would not be shocked, and should not touch the inductorium when the light dimmed or else she would be shocked. This condition was used to assess the arousal effect of an electrified inductorium without shock to the performer.

Twenty-four subjects were assigned to each condition so that each condition contained an equal number of males and females.

A female confederate acted as the performer in this experiment. In addition to the different instructions for the various conditions, there were some procedural changes. In Experiment I, an attempt was made to use a "blind" performer by having the critical instructions for the different conditions read to the observer before the performer entered the laboratory. This procedure was abandoned in this experiment for two reasons: the performer in Experiment I frequently could identify the condition being used by the kinds of overt responses the observer made (e.g., smiling, looking away when the performer jerked, apparent boredom, etc.), and the performer in this experiment had to know what condition she was in so that she could behave appropriately.

The criteria for scoring GSRs were the same as in Experiment I, except that the pen

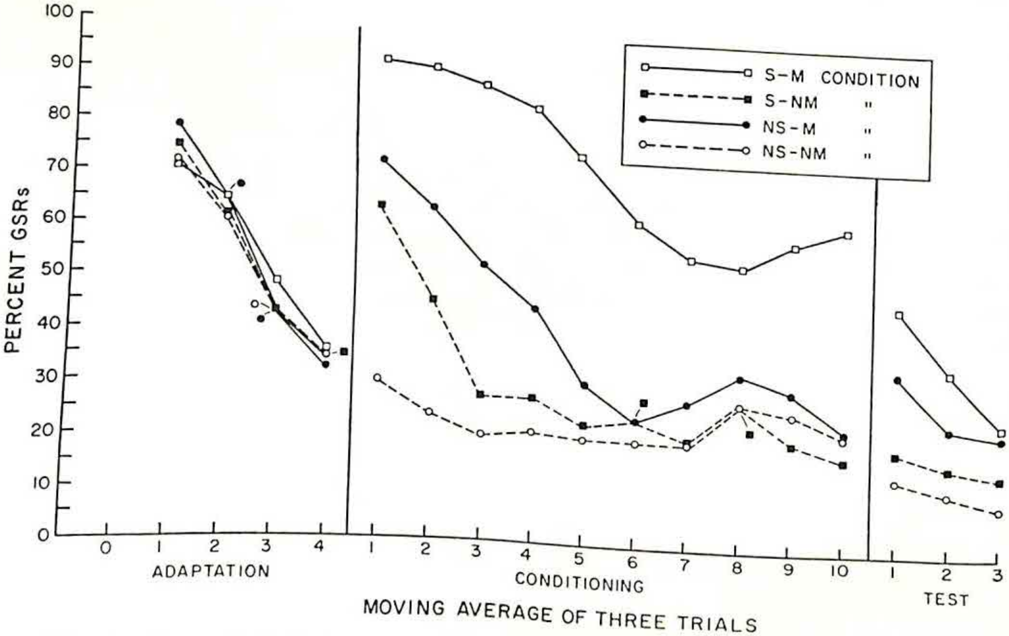


FIG. 2. Percentage of observers giving GSRs during adaptation, conditioning, and test trials (Experiment II).

movement in the direction of lowered resistance need only have been greater than the thickness of the pen line (rather than twice the thickness). This change represented an attempt to increase the sensitivity of the GSR measure, particularly with regard to CRs, which generally occurred relatively infrequently in Experiment I. Scoring was done independently, and without identification of the condition to which individual observers were assigned. The scorers agreed on 96% of the trials concerning the occurrence of GSRs; disagreements were resolved by reexamination and discussion by the scorers.

No differences in observer GSRs were observed between conditions during the adaptation period, as may be seen in Figure 2.

Table 2 summarizes the analysis of variance performed on the total scores for observers in the different conditions over the 10 conditioning trials.

Since the purpose of this study was to determine whether the differential effect of shock instructions found in Experiment I would occur without performer arm movement, the Shock × Movement interaction in the analysis of variance is particularly relevant. This interaction indicates that the differential effect of shock instructions depends upon whether or not performer arm movements are present. There is a significant difference between the frequency of GSRs given by S-M observers and the frequency of GSRs given by observers in the NS-M condition ($t = 4.33$, $df = 46$, $p < .01$); this difference is not significant between the observers in the S-NM and NS-NM conditions.

The main effect for Movement reveals that performer arm movements are effective elicitors of the GSR, regardless of shock instructions. The t between S-M and S-NM observers is 5.89, $df = 46$, $p < .01$; the t between the NS-M and NS-NM observers is 2.92, $df = 46$, $p < .01$.

The main effect for Shock is primarily due to the difference between the S-M and NS-M observers, the difference between the S-NM and NS-NM observers being small and not significant.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF GSR INSTIGATION SCORES

Source	df	MS	F
Shock instructions	1	94.01	16.79**
Performer movement	1	219.01	39.11**
Sex of observer	1	4.60	—
Shock × Movement	1	38.76	6.92*
Shock × Sex	1	3.01	—
Movement × Sex	1	7.58	1.36
Shock × Movement × Sex	1	2.99	—
Residual	88	5.60	

* $p < .025$.
** $p < .001$.

The t tests reported above show that observers in the S-M condition differ from observers in the NS-M and S-NM conditions. The S-M observers also differ from NS-NM observers ($t = 7.35$, $df = 46$, $p < .01$).

The differences between conditions over conditioning trials may be seen in Figure 2.

The results of Experiment I suggested instigation was not solely a function of the performer's arm movement. This finding is strongly supported by the results of this experiment; a very reliable difference was found between the S-M and the NS-M conditions. In addition, this study reveals that the instructions concerning shock to the performer (S-NM condition), or that electricity is being passed to the inductorium (NS-NM condition) are not sufficient in themselves to account for the frequency of observer GSRs in the S-M condition. It is apparent from these data that frequency of instigation depends upon the instructions concerning shock and the occurrence of the performer's response.

Figure 2 also presents the test trial data for observers in four conditions. An analysis of variance of the test trials scores revealed no reliable differences in conditioning. The order of conditions, based upon the magnitude of the means for CRs, is in line with the order of conditions based upon the means for instigations. The frequency of CRs tends to be greatest in the S-M condition, although the difference between S-M observers and observers in the other conditions was not great enough to yield statistically significant results.

Although reliable differences were found between the conditioning results from the S-M condition and the NS-M condition in Experiment I, the differences between responsiveness on test trials for these conditions in this study were not reliable, but in the same direction.

A content analysis of the post-experiment questionnaire data attempted to answer two questions: Did the observer describe the performer's actions in accordance with the instructions he received? Upon questioning, did the observer report that he expected a shock during the last part of the experiment?

In the S-M condition, one observer reported that he did not think the performer was being shocked, while this report was given by eight observers in the S-NM condition. Judging from their reports, a large number of observers in the S-NM condition believed that the performer was receiving a very mild shock or no shock at all; this may account for the relatively low frequency of instigations in this condition. Many S-NM observers did not perceive that the performer was being shocked or made uncomfortable. Five observers in the NS-NM condition reported that they did not think the performer would be shocked if she touched the inductorium. The majority of observers, however, described the conditions of the experiment in accordance with the instructions given, the one major exception being the S-NM observers.

Most of the observers who reported that they expected to receive a shock during the last part of the experiment, apparently expected the shock on the first trial or two, and when it did not happen, they no longer expected to be shocked.

When the mean GSRs from observers who expected shock are compared with the general means for the relevant conditions, it appears that all means are essentially equal to or less than the general mean for the particular condition—except for the instigation mean of S-M observers. Of the two observers in the S-M condition who expected shock, one gave GSRs on 8 out of 10 conditioning trials, which is equal to the median number of responses given by observers in the S-M condition; the other observer gave GSRs on all 10 trials, which is equal to the score of one-third of the observers in the S-M condition. Of the 24 observers in each condition, 2 S-M observers, 3 S-NM observers, 5 NS-M observers, and 5 NS-NM observers reported some expectation of receiving shock. Consequently, even though a small number of observers in each condition reported that they expected to be shocked, there is nothing in their data to suggest that this is a significant factor in their GSR responsiveness.

While there is general agreement between these findings and the results of Experiment I, the differences between the S-M and NS-M conditions with respect to instigation and conditioning were not consistently as large in both studies. When the probabilities of the observed differences are considered together, there is good support for the phenomenon of conditioning through vicarious instigation. Yet, it appears that some refinements in measuring techniques might improve the reliabilities of the differences.

EXPERIMENT III

There is some possibility that the instructions and performer reactions resensitized the observer's response to the buzzer during the second part of the experiment, so that the measures of instigation and conditioning were confounded with a resensitized response to the buzzer, per se. Furthermore, the use of only three test trials may have been insensitive to some of the conditioning that occurred. For these reasons, the measurement technique was changed to conform to the recommendations of Stewart, Stern, Winokur, and Fredman (1961), in which the latency of the GSR was used to identify response measures. By taking into account the latency of the GSR to a stimulus, these authors found it possible to obtain separate measures of responses to the CS and UCS, and to identify conditioned anticipatory responses, for each trial. With this technique, therefore, it is possible to have each presentation of the stimuli serve as a test trial for conditioning, and simultaneously obtain measures of responses to the CS and to the UCS. The purpose of this experiment was to determine the utility of the latency technique for conditioning through vicarious instigation. Only the S-M and NS-M conditions were examined in this experiment, using the same instructional procedures as reported for Experiment II.

Some procedural changes with regard to presentation of stimuli were introduced to accommodate the latency technique. During adaptation, only the buzzer was presented for 6.5 seconds. The intertrial interval was varied at random (during adaptation and conditioning) from 20 to 60 seconds. During conditioning, the buzzer was presented for the same length of time as in adaptation. The dimming of the light for 1 second followed the termination of the buzzer by .5 second. As usual, the performer jerked her arm when the light dimmed.

For scoring purposes, the stimulus interval was divided into three 4-second units.[8] The first 4-second unit represented the response to the buzzer; the last 4-second unit, measured from the end of the light dimming period, represented the response to the performer's movement; the second 4-second unit represented a conditioned anticipatory response.

The criteria for identifying GSRs was the same as used in Experiment II, except that the pen movement in the direction of lowered resistance had to occur within one of the 4-second intervals, and the movement had to indicate an accelerating change rather than a decelerating one; this latter criterion seemed necessary to distinguish GSRs initiated in one scoring period from one initiated in the next period, on a given trial. Again, scoring of the records was done independently, without identification as to which condition the observer had been assigned. The scorers agreed on 93% of the cases, and resolved their disagreements by discussion.

The performer was a female in this experiment, but not the same person who was used in Experiment II.

Sixteen subjects were assigned to each condition so that each condition contained an equal number of males and females.

The adaptation for the two conditions was quite similar. Figure 3 shows the response frequencies to the buzzer, the second, and the third scoring units for both conditions over the first four adaptation trials. Since only the buzzer was presented during adaptation, responses occurring during the second and third scoring periods are assumed to be spontaneous.

When responses to the buzzer during conditioning are compared to the buzzer responses occurring during adaptation (see Figure 3), it is apparent that observers are resensitized in the conditioning period. An analysis of variance based upon total responses

---

[8] The 4-second unit was used in Experiments I and II for identifying GSRs to the stimuli. It is also in agreement with the general findings of Stewart, Stern, Winokur, and Fredman (1961, p. 63).

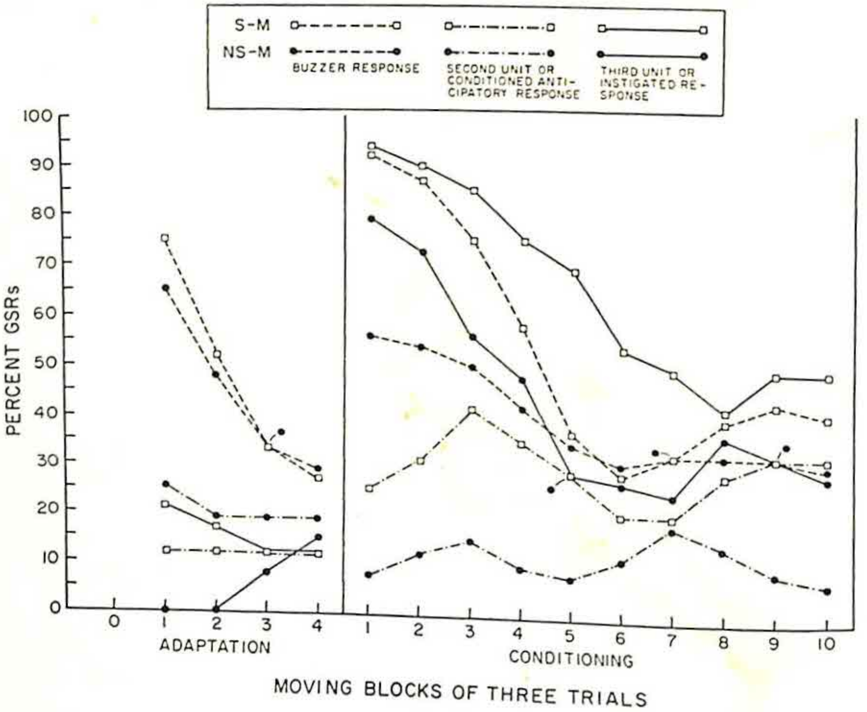


FIG. 3. Percentage of observers giving GSRs during adaptation and conditioning (Experiment III).

to the buzzer during conditioning revealed no reliable differences between the two conditions.

The curves presented in Figure 3 for instigated responses have the same general shape as the curves for the comparable conditions in the two previous studies; S-M observers are more responsive than NS-M observers. The analysis of variance based upon total instigations for each observer shows that this difference is reliable (see Table 3).

S-M observers gave more conditioned anticipatory responses than NS-M observers, as may be seen in Figure 3. The analysis of variance, summarized in Table 4, shows that the difference in conditioned anticipatory response totals for observers in the two conditions is reliable.

In this experiment, the experimenter asked each observer at the end of the experiment if he expected to receive a shock. Only one observer in each condition reported that he expected shock.

These findings provide additional support for the phenomenon of conditioning through vicarious instigation. The general level of observer instigation was similar to that found for comparable conditions in Experiment II, but a general increase in the number of conditioned GSRs was noted in Experiment III, where anticipatory re-

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE FOR INSTIGATIONS

Source	*df*	*MS*	*F*
Shock instructions	1	40.50	5.36*
Sex of observer	1	8.00	1.06
Shock × Sex	1	0.00	
Residual	28	7.77	

* $p < .05$.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE FOR CONDITIONED ANTICIPATORY RESPONSES

Source	*df*	*MS*	*F*
Shock instructions	1	26.28	9.16*
Sex of observer	1	2.53	—
Shock × Sex	1	1.53	—
Residual	28	2.87	

* $p < .01$.

sponses were examined on all 10 trials by the latency criterion.

DISCUSSION

Considering all three studies together, there is a good deal of consistency in the findings with regard to the S-M and NS-M conditions. Although the differences between these conditions for instigations and conditioned responses was not consistently reliable in Experiments I and II, the trends in the differences were always in the same direction. With particular reference to the CR data, the finding in Experiment III that observers in both conditions are resensitized to the buzzer suggests that this factor may have tended to mitigate the reliability of the differences.

The reader may be concerned with the possibility that the differences between conditions may be influenced by different base levels of resistance, produced by the different treatments. The base levels of resistance (in ohms) were generally lower for S-M observers than for observers in the control conditions in these studies, although frequency of GSRs was greatest in the S-M condition. Since there is a positive relationship between GSR size and base level of resistance (Haggard, 1945), this would suggest that the effect of base level differences tended to decrease the differences in frequency of GSRs between the S-M and control conditions, according to the GSR criteria used. Consequently, if base level differences did influence the results of these studies, it was in the direction of minimizing vicarious instigation effects.

It should be noted that the present experimental situation *underplays* the instigation of the performer. This was done in order to keep the S-M and NS-M conditions plausible and yet comparable. Presumably, the level of instigation (and consequently, conditioning) would be higher if the performer's behavior was changed so as to provide greater evidence of pain. Actually, the present level of instigation and conditioning is high enough to provide evidence for the phenomenon, and yet low enough to provide a basis for testing the facilitating effects of a variety of variables.

In all of the data, observer GSRs tend to occur less frequently during the last half of the conditioning period than the first half. The explanation for this change is not entirely clear. One possible explanation is that GSR adaptation occurs because the observer becomes habituated to the specific situation as a result of repeated presentations of the stimuli, as Sears (1933) has shown. In addition, relatively short intertrial intervals tend to produce rapid GSR adaptation to repeated stimuli (Coombs, 1938). It also is possible that observers pay less attention to the stimuli as the experiment progresses. For example, some observers in the S-M condition reported that they could not bear to watch the performer being shocked and looked away after the first few trials; observers in the NS-M condition may have become bored with, and less attentive to, the repetitious movement.

It might also be argued that vicarious instigation phenomena have a transitory quality, the effect diminishing with successive exposures to the performer's situation as the observer begins to exercise some internal controls over his ERs. An adult audience witnessing a movie, for example, may be emotionally aroused as they watch their favorite star fall off a cliff for the first time; but their level of arousal will probably be lowered after witnessing the same scene over and over again. Individuals can exercise control over their emotions when they can anticipate the occurrence of an emotionally arousing situation; although, considering the number of times a child can sit through the same movie with apparently undiminished enthusiasm, it would seem that children are less capable than adults in exercising such control.

Since it cannot be determined, from the data reported, whether the adaptation effects noted in these studies are a function of the characteristics of the GSR or a function of the characteristics of vicarious instigation phenomena generally, it would seem that a different measure of vicarious instigation is needed in order to avoid this confounding.

At the same time, some measure is needed to identify the nature of the observer's ER. While the reader may be tempted to think of these experiments as reflecting some sort of observer *empathy,* there is no basis for this interpretation. Some observers may be empathizing and some observers may be enjoying the performer's apparent pain. Research is presently underway which employs the evaluative scales of the Semantic Differential for this purpose. It may be possible to measure conditioning (as Staats & Staats, 1957, have done), with the performer's UER as the UCS for the observer. If this research is successful, it not only should provide a basis for determining how attitudes and values may be conditioned through vicarious instigation, but it should be possible to examine, independently of the GSR, whether adaptation of vicariously instigated responses occurs.

The general experimental situation used in these studies appears to be a useful one for measuring conditioning through vicarious instigation. It should be noted that the overall findings lend support to the concept of vicarious instigation as defined in the introduction to this paper: the instigation and conditioning of the observer's GSR cannot be attributed to the performer's movement, per se, or to the electrified inductorium, per se, or to the instructions concerning shock, per se. It is the apparent shock of the performer (and by inference, the performer's UER) which is the effective instigating stimulus for observer response.

Vicarious Instigation and Conditioned Emotional Response

Vicarious instigation is certainly not a new problem in social psychology. The phenomenon has been identified many times by other terminology. McDougall (1908) referred to the phenomenon as the "sympathetic induction of emotions." Floyd H. Allport (1924) reacted strongly against this instinct doctrine and sought to bring the phenomenon within the scope of conditioning theory.

> A theory far more plausible than that of instinctive induction of emotion . . . is the principle of *conditioned emotional response.* It may be illustrated by the panic which seizes all the persons in a throng when a few of them show signs of terror. Granting that the true cause for alarm has been seen by only the original few, we have here a case of fear aroused by the process of sympathy. The explanation, according to the present theory, is as follows: we have been previously terrified in company with others and so have our fear emotion transferred to characteristic attendant stimuli, such as cries and visible expressions of the emotion in those about us. We now react at once to

> the sight of fear in others by a fear response of our own. Here the conception of sympathetic induction loses its force. We fear not merely because we see the expression of fear in others; but because we have learned to read these expressions as signs that *there really is something to be afraid of.* It is not fear induced from others that we experience, but *our own* fear of dangerous situations which has been conditioned by social stimuli (p. 235).

Allport's (1924) analysis argues against vicarious instigation, as defined in this paper; i.e., the expressed emotions of others are simply discriminative stimuli for subsequent presentation of the UCS to the observer. His analysis may hold for *some* situations, but it does not seem applicable to the experiments just reported. Observers were reassured that they would not be shocked, and therefore had nothing to fear for themselves; and most observers reported that they were not concerned for themselves. Furthermore, the results of the experiments show that the arousal effects are greatest for S-M observers, when the performer's situation is made to appear most unpleasant; but the number of S-M observers who apparently expected to be shocked is equal to or less than the number of observers who expected shock in the less arousing conditions.

While these arguments are not conclusive, they suggest that vicarious instigation is either a special case of conditioned emotional responding or that vicarious instigation depends upon a different learning principle. The resolution of this issue is not imminent, since further clarification of the nature of conditioned emotional responses and vicarious instigation is required before suitable arguments can be formulated.

Vicarious Instigation in Relation to Experimental Operations

The use of verbal instructions to manipulate the independent variables, and a post-experiment questionnaire to check the subject's acceptance of the instructions, limits the application of the designs used for experimental analysis of vicarious instigation to certain classes of human subjects. It is clear that these experimental operations are not appropriate for work with lower animals, young children, or with mental retardates who have little verbal facility.

One possible solution to this problem is to substitute specific training histories for the instructions. Observers first may be given direct experience with the condition that the performer will undergo, and then be exposed to a performer in that condition. Subsequently, the observer may be tested to determine if his behavior was affected by exposure to the performer. Although it seems feasible to provide different training histories for each experimental and control condition used in studies of vicarious instigation, experimental evidence is needed to demonstrate the suitability of the training technique for investigation of vicarious instigation phenomena.

To some extent, the present definition of vicarious instigation has a certain conceptual inadequacy. In order to identify conditioning through vicarious instigation, it is necessary to show that sources of pseudovicarious instigation do not account for the behavior in question. From an experimental viewpoint, the presence of a particular factor is more easily demonstrated than its absence. Until a more positive statement can be formulated concerning the necessary and sufficient conditions for vicarious instigation, it may be anticipated that the present definitions will tax the experimenter's ingenuity for setting up experimental situations which eliminate sources of the pseudoeffects mentioned previously. In spite of this inadequacy, the conceptual framework has experimental value.

CONCLUSION

This paper offers a conceptualization of conditioning through vicarious instigation and provides some experimental support for the phenomenon. The importance of the problem is made evident by the paucity of experimental evidence concerning the role of the phenomena of empathy, envy, and sadism in the acquisition of interpersonal behavior. The analysis of pseudovicarious instigation dramatizes the inadequacy of anecdotal evidence in this area.

No systematic attention has been given here to the personality and social determinants of vicarious instigation, or to the consequences of vicariously instigated conditioning for social interaction. Heider (1958), for example, identifies a wide range of common sense principles which could be incorporated into the conditioning models presented in this paper. One of the virtues of these models is that they permit the perceptually oriented theorist to apply his principles in a conditioning framework and allow the S-R learning theorist to test his principles with a wider range of stimuli.

REFERENCES

ALLPORT, F. H. *Social psychology.* New York: Houghton Mifflin, 1924.

ALLPORT, G. W. Attitudes. In C. Murchison (Ed.), *A handbook of social psychology.* Worcester, Mass.: Clark Univer. Press, 1935. Pp. 798–844.

ASCH, S. B. *Social psychology.* New York: Prentice-Hall, 1952.

BARNETT, P. E., & BENEDETTI, D. T. A study in "vicarious conditioning." Paper read at Rocky Mountain Psychological Association, Glenwood Spring, May 1960.

BERGER, S. M. Incidental learning through vicarious reinforcement. *Psychol Rep.,* 1961, **9**, 477–491.

BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.,* 1960, **57**, 49–64.

CAMPBELL, D. T. Conformity in psychology's theories of acquired behavioral dispositions. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation.* New York: Harper, 1961. Pp. 101–142.

CHURCH, R. M. Emotional reactions of rats to the pain of others. *J. comp. physiol. Psychol.,* 1959, **52**, 132–134.

COOMBS, C. H. Adaptation of the galvanic response to auditory stimuli. *J. exp. Psychol.,* 1938, **22**, 244–268.

HAGGARD, E. A. Experimental studies in affective processes: II. On the quantification and evaluation of "measured" changes in skin resistance. *J. exp. Psychol.,* 1945, **35**, 46–56.

HANER, C. F., & WHITNEY, E. R. Empathic conditioning and its relation to anxiety level. *Amer. Psychologist,* 1960, **15**, 493. (Abstract)

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* (2nd ed. rev. by G. A. Kimble) New York: Appleton-Century-Crofts, 1961.

HILL, W. F. Learning theory and the acquisition of values. *Psychol. Rev.,* 1960, **67**, 317–331.

McDOUGALL, W. *An introduction to social psychology.* London: Methuen, 1908.

MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.,* 1948, **38**, 89–101.

MURPHY, J. V., MILLER, R. E., & MIRSKY, I. A. Interanimal conditioning in the monkey. *J. comp. physiol. Psychol.,* 1955, **48**, 211–214.

SEARS, R. Psychogalvanic responses in arithmetical work. *Arch. psychol., N. Y.,* 1933, No. 155.

SECHREST, L. Vicarious reinforcement of responses. *Amer. Psychologist,* 1961, **16**, 356. (Abstract)

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.,* 1957, **54**, 74–80.

STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.,* 1961, **68**, 60–67.

(Received March 13, 1961)

Psychological Review
1962, Vol. 69, No. 5, 467–476

CONDITIONAL–OUTCOME CHOICE BEHAVIOR IN RATS[1]

FRANK A. LOGAN

Yale University

A conditional-outcome choice has the property that, for at least one of the alternatives, the outcome itself depends upon the subject's behavior subsequent to the choice. Because this type of situation is somewhat novel, it may be useful to begin with an example.

Figure 1 shows the floorplan of a maze in which a rat can run from the start (S) around one of two perimeter paths to the goal (G). If he turns to the right after leaving the start box, his path takes two additional right turns and he then arrives at a second choice point, shaped somewhat like a **Y** maze and exposing two swinging doors. One of these doors is white and the other black, and reward at the goal is given if the rat selects (say) the white door regardless of its position. That is to say, the second choice point contains a conventional visual stimulus discrimination problem. Alternatively, if the rat turns to the left after leaving the start, his path takes two additional left turns and, at the location corresponding to where in the other wing he encounters a discrimination problem, here one of the two guillotine doors is closed forcing him, half the time to the left, half the time to the right, through a grey swinging door into the goal. Typically, he is fed in this case with the frequency that he is earning reward by his discrimination performance in the other wing.

The important thing to recognize is that, for the first alternative described, the outcome cannot be specified in advance. Whether or not the subject gets reward depends upon what he does, in this case, upon whether he goes through the correct door. It is this feature that the subject has control over the outcome that characterizes conditional-outcome choices.

This type of situation has considerable interest in its own right. As a single but familiar example, a person selecting a profession cannot just consider various jobs with associated probabilities of payoffs because the payoff depends upon how he performs in his chosen field. How much one earns as a scientist, a doctor, or even a ditch-digger depends in part upon how good he is at that job. The outcome, in short, is conditional upon performance.

But beyond this intrinsic interest, conditional-outcome choices have an important relationship to the problem of response definition. The discussion of this relationship in the next section will also show the logic of the researches to be reported in subsequent sections.

Conditional-Outcome Choices and the Problem of Response Definition

Fundamentally, the problem of response definition is one of deciding upon a rule, or set of rules, stating the

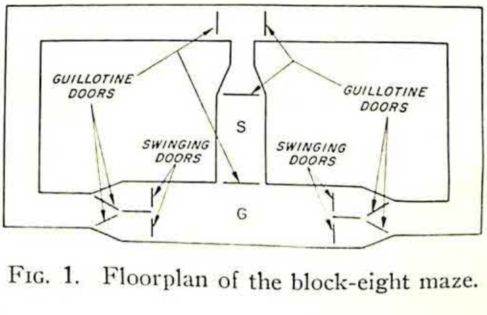


FIG. 1. Floorplan of the block-eight maze.

---

[1] Supported in part by the National Science Foundation.

conditions under which behaviors that differ in some way may nevertheless be aggregated into a single response class. All familiar response terms are aggregates of one kind or another. For example, there are a great many ways that a rat can get from one end of an alley to the other but these are all lumped together indiscriminately in the expression, the rat "ran down the alley."

The conventional solution of this problem seems to be that responses can be aggregated so long as they are not differentially reinforced (see, e.g., Guthrie, 1940; Hull, 1943). Thus, no distinctions are made among the various ways of running down an alley because the reward that is given at the end of the alley typically does not depend upon how the rat got there. This approach is schematized for a bar-pressing situation in Figure 2.

In the upper left box are identified four responses that can occur in a two-bar Skinner box: the rat can press Bar A with the left or right paw, or can press Bar B with the left or right paw. Of course, each of these is already an aggregate of different postures, different forces, and so forth, but they will suffice for the purposes of discussion. Now the typical procedure is to reinforce the rat differentially with respect to which bar he presses but not to reinforce differentially which paw he uses. Such a condition is indicated by the wavy arrows showing that the rat gets 8 pellets for each of the Bar A responses and 2 pellets for each of the Bar B responses. According to the rule that responses that are not differentially reinforced can be aggregated, the description can be simplified to that shown in the upper right box. In this case, information about which paw the rat used is completely suppressed and one simply refers to Bar A responses and Bar B responses.

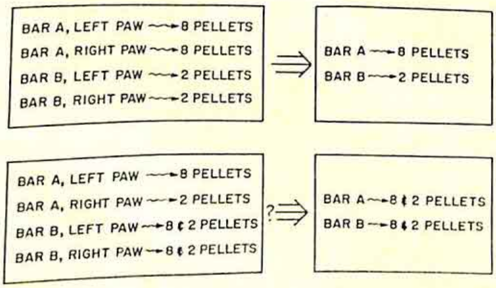

BAR A, LEFT PAW → 8 PELLETS
BAR A, RIGHT PAW → 8 PELLETS
BAR B, LEFT PAW → 2 PELLETS
BAR B, RIGHT PAW → 2 PELLETS
⇒
BAR A → 8 PELLETS
BAR B → 2 PELLETS
BAR A, LEFT PAW → 8 PELLETS
BAR A, RIGHT PAW → 2 PELLETS
BAR B, LEFT PAW → 8 & 2 PELLETS
BAR B, RIGHT PAW → 8 & 2 PELLETS
?⇒
BAR A → 8 & 2 PELLETS
BAR B → 8 & 2 PELLETS

FIG. 2. Schematization of the responses in a two-bar Skinner box for an unconditional-outcome choice between the bars (top boxes) and a conditional-outcome choice between the bars (lower boxes).

The justification for this procedure is the same one that applies to the formation of any meaningful scientific concept, namely, that responses defined in this way enter into reasonably simple and consistent lawful relationships with other concepts (see, e.g., Spence, 1948). The substantial literature on choice behavior provides ample evidence of the usefulness of this procedure.

Suppose, however, that the reinforcement conditions were arranged as shown in the lower left box. In this case, the rat is differentially reinforced with respect to which paw he uses on Bar A, getting 8 pellets for left-paw responses and 2 pellets for right-paw responses. On Bar B, however, the reward is sometimes 8 and sometimes 2 pellets but this variation is not conditional upon which paw the rat uses.

Specifically, the question being raised is this: can the two Bar A responses be aggregated even though they are differentially reinforced? Can we suppress information about which paw the rat uses, and even ignore the fact that the reward is conditional upon which paw he uses? In short, can we, for purposes of predicting choice, refer to Bar A responses as shown in the lower right box? If so, then *that* response receives varied reinforcement, sometimes 8 and sometimes 2 pellets.

Again, the question of whether this step can be taken means, will a response defined in this way act like familiar responses that only aggregate nondifferentially reinforced behaviors? Will such a response obey the same laws and principles as determined under conventional conditions? One way of asking this question is to arrange the reward on Bar B so that the variation between 8 and 2 pellets there exactly matches that being received on Bar A as a result of trial-to-trial oscillation in the rat's behavior. If the aggregation is legitimate, then performance under these conditions can be predicted from familiar principles because the reward is not differential with respect to which bar the rat presses and he should, therefore, be indifferent between them when given a choice.

However, if the aggregation is not appropriate, if the familiar principles do not apply, or if new principles are required by conditional-outcome choices, then a preference might appear between alternatives that are otherwise equally reinforced. Specifically, for example, there might be something positive to a rat about having the reward under his control, in which case he would prefer an alternative such as Bar A where what he gets depends upon what he does. Or, with equal a priori possibility, such a condition could be aversive, in which case the rat would prefer not to have to make a differentiation on the way to the reward and so would choose an alternative such as Bar B. The following researches have been aimed at evaluating these possibilities.

METHOD

Subjects. Naive male rats were obtained from a variety of commercial sources at 90–120 days of age. They were housed in individual cages and were fed 12 grams of Purina laboratory chow daily 5–10 minutes after being run. Reward consisted of four 45-milligram Noyes pellets.

Apparatus. The studies of choice behavior when the outcome was conditional upon running speed were run in the double-alley maze that has been described in detail elsewhere (Logan, 1960). Briefly, it consists of a grey start box leading to two parallel alleys, one of which is white and the other black, both 4-feet long and equipped so that food reward can be delivered automatically according to a wide range of possible conditions.

The studies of choice behavior when the outcome was conditional upon a visual discrimination or a position response were run in the block-eight maze shown previously as Figure 1. This maze is grey throughout and is illuminated by six 7-watt bulbs spaced symmetrically around the maze and 18 inches above it. Food reward is given manually when appropriate by pulling a rod outside the maze that uncovers the food cup.

Procedure. Trials were run in blocks of four or six, one trial in each block being a free choice trial and the remaining trials being forced so as to insure that the rat went equally often in each direction. The sequence of forcing was arranged so as to avoid learnable sequences and to counterbalance alternation tendencies. The number of such blocks of trials run in a day varied from one to six in the various studies, and the intertrial interval was in some cases only the few seconds necessary to move the rat from the goal to the start box and in other cases was the time required to run the other concurrent rats. On every occasion that the same problem has been studied using procedures differing in any of these ways, the results have been closely comparable and there is no reason to believe that the results are peculiar to a particular aspect of the procedure.

One detail of the procedure that may be especially important concerns the technique for matching frequency of reward in the two alternatives. Half of the trials in each block were to each alternative and the rewards given in the unconditional alternative in any block of trials duplicated the rewards that the subject had earned in the conditional alternative on the preceding block of trials. Hence, there was a slight unavoidable lag in matching the rewards.

Outcome Conditional upon Running Speed

One way that the reward can be made conditional upon the rat's be-ward conditional upon running speed, the effect is not very large.

Correlated reinforcement versus matched frequency and interval of reward. Our final attack upon this problem attempted to equate both the frequency and the interval of reinforcement in the two alleys. To do this, the problem in the correlated alley was changed somewhat, the rats in this case being required to run down to the entrance to the goal section of the alley and to wait there for at least 2 seconds before proceeding to the goal. If they failed to spend at least 2 seconds in that foot of the alley immediately preceding the goal, they were not rewarded. This restriction on the correct response was made so that the guillotine door in the alternative alley could be closed to detain the rats at the same distance from the goal and for a length of time matched to the time they detained themselves in the correlated alley. Since the frequency of reward was also matched, there is no apparent bias in the reward conditions favoring one alley over the other and there should therefore be no preference between them.

Four rats have been run under these conditions and, during the first 1,024 trials, all four developed and maintained 100% preference for the correlated alley, for the alley in which they timed their delay at the entrance to the goal section of the alley. However, when the conditions were reversed for three of these animals, only two reversed their preference within the next 1,024 trials, the third continuing to choose the same direction which was now the uncorrelated alley. Hence, there is some evidence for a positive aspect to an alternative in which the reward is conditional upon speed, but the effect is either not very large or not permanent.

Outcome Conditional upon a Discrimination Problem

A second way that the reward can be made conditional upon the rat's behavior is to use a visual discrimination problem. This situation has already been described, it being one in which, if the rat goes in one direction in the maze shown in Figure 1, he encounters a second choice point at which he must choose between a black and a white door, and if he goes in the other direction, he is forced through a grey door. Frequency of reward is matched in the two wings and hence familiar principles would lead one to expect the rat to be indifferent between them.

Black-white discrimination versus matched frequency of reward. Fourteen rats have been run with a black-white discrimination problem in one wing of the block-eight maze, and their combined performance is shown in Figure 5. There were wide individual differences in rate of learning the discrimination problem but the graph nevertheless correctly reflects the relevant features to be seen by examining all the individual curves. As the discrimina-

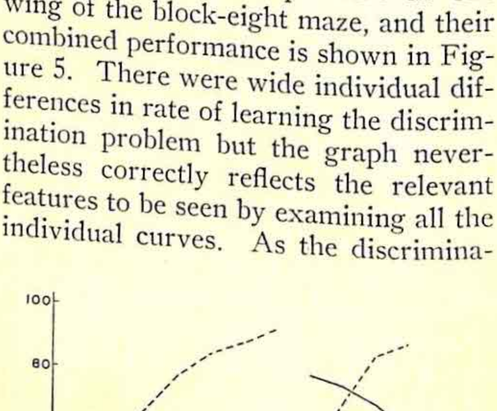


FIG. 5. Combined performance of 14 rats that encountered a black-white dsicrimination in one wing of the block-eight maze and were given a matched frequency of reward in the other wing. (The break in the curves indicates where the discrimination problem was placed in the opposite wing.)

tion problem was initially solved, a preference for the nondiscrimination wing appeared, only two rats choosing the discrimination wing on more than half the last 160 trials. But there was not a clear reversal in preference when the location of the discrimination problem was reversed. Hence, there is some evidence for a negative aspect to an alternative in which reward is conditional upon a visual discrimination, but the effect is either not very large or not permanent.

Dark-light discrimination versus matched frequency of reward. The direction of the effect with a discrimination problem was opposite to that obtained when reward was conditional upon running speed. There are several differences other than the nature of the task that might account for this discrepancy. Most obviously, there is a difference in difficulty because a rat on a correlated reinforcement condition in the alley initially gets reward very infrequently and never attains more than 75–80% correct, whereas a rat on a discrimination problem begins with 50% reward and ultimately attains essentially perfect performance.

In an attempt to explore this possibility, the discrimination problem in its reversed position was made more difficult for eight of the rats by decreasing the difference between the brightnesses of the two doors, generally by converging on an intermediate grey. This form of transposition was very good, the rats by-and-large continuing to choose the lighter or darker door depending upon the way they had originally been trained, but the difficulty level was in this way effectively increased as evidenced by a decrease to about 85% success. There was no observable effect upon choice as between the discrimination and nondiscrimination wings of the maze after 1,280 trials under these more difficult conditions, 59% of the choices in the last 160 trials being to the discrimination side.

Black-white discrimination versus dark-light discrimination. As a further test of the effect of difficulty of the task, a discrimination problem was then placed in both wings of the maze, one being the simple black-white problem and the other, a more difficult discrimination between two greys. Darker (or lighter) was correct on both sides. In order to avoid a preference developing between these alternatives because of a difference in reward based on a difference in percentage success, the frequency of reward in the simple discrimination was restricted to the same level that the rat was obtaining on the more difficult problem. There was no evidence of a preference after 1,280 trials under these conditions. The rats all continued to choose the direction they had been preferring during the previous phase of the experiment, which for some rats meant choosing the more difficult wing and for others, choosing the easier wing. Hence, problem difficulty, at least for experienced animals with matched frequency of reward, is not a critical variable in conditional-outcome alternatives.

Outcome Conditional upon a Position Response

A third type of situation in which the general problem of conditional-outcome choices can be tested involves the same maze as used for visual discrimination problems, but requires a position response at the second choice point. In this case, the swinging doors are all grey and in one wing of the maze, reward is conditional upon the rat going through (say) the left door. In the other wing of the maze, the turn is forced and reward frequency is matched to the percentage correct in the conditional alley.

One difficulty with this test of choice behavior when one alternative is conditional upon response selection is that the same type of response, namely, a position response, is also used to determine the rat's preference. If there is any anticipatory response tendency, it could contaminate the measure of preference; e.g., a preference for a left turn at the first choice point might appear because a left turn is correct at the second choice point rather than because that direction leads to (or away from) the position problem. A similar complication would have arisen in the previous cases had a stimulus discrimination or response speed differentiation been used as the indicator response. For example, a rat could encounter a black-white discrimination problem at a second choice point if he chooses white at the first choice point and not if he chooses black there, in which case the differential reinforcement of brightness at the second choice point might transfer to the first. It may be useful later to repeat these studies using different indicator responses.

However, there are several reasons for believing that this problem is probably not serious in the present instance. There is evidence (Logan, 1961) that learning is quite specific to its location in this apparatus, and unpublished research by students failed to detect an anticipatory response tendency. Furthermore, detailed examination of the present data did not reveal any consistent relationship between preference at the first choice point and the correct turn at the second.

Seven rats have been run on this type of problem and Figure 6 shows the results for four of them for whom location and correct position were counterbalanced. Only one rat showed a strong preference for the conditional-outcome alternative but the overall picture leans somewhat in that direction.

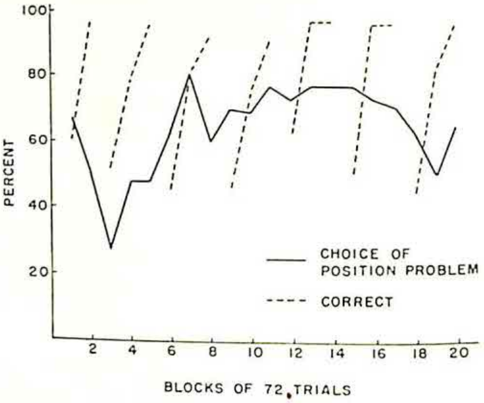


FIG. 6. Combined performance of four rats that encountered a position problem in one wing of the block-eight maze and were given a matched frequency of reward in the other wing. (Each break in the dashed curve indicates where the correct turn at the position problem was reversed.)

Hence, there may be a slight positive aspect to an alternative in which reward is conditional upon a position response.

Discrimination Problem versus Position Problem

Since the evidence suggested the possibility of something positive about a position problem and something negative about a discrimination problem, when each was compared with an unconditional outcome, it seemed reasonable that these effects might appear more clearly if they were pitted against each other. Toward this end, a black-white discrimination problem was placed in one wing of the block-eight maze and a position problem was placed in the other wing.

These conditions could not be used from the beginning of training because the two wings would be differentially reinforced. To avoid this difficulty, the rats were run through three phases preceding the testing phase. The first phase entailed running for 864 trials in just one wing to a discrimination problem or, for half the rats, a position problem. The second phase entailed running for 864 trials in the opposite wing to the other problem. Because the position problem could be learned more quickly than the discrimination problem, the rats were given several reversal series in the position wing so that they received about the same number of reinforcements in each wing during the first two phases. The third phase entailed integrating the two problems, the rats being given an irregular series of 192 forced trials half in each wing with the appropriate problem in place. Four of the six rats made this transition to the integrated problem with only a few errors, but the other two had to be returned to a previous phase and then brought back to the integrated problem.

Hence, by the time of the fourth phase during which the customary sequence of free and forced trials was followed, the rats had solved both problems, gone each direction equally often, received comparable reinforcement frequencies in each wing, and currently were performing essentially perfectly on both problems. There is thus no basis in familiar principles to predict a preference.

None of the six rats showed a strong preference on the free trials during the fourth phase. Indeed, they were conspicuously indifferent, overall percentage choice of the position problem being 55%. Hence, there is no evidence that the nature of the problem makes a difference in conditional-outcome choices so long as frequency of reward is equated.

DISCUSSION

The reader may wish to emphasize those few instances in which we have obtained preferences that do not appear to follow from familiar principles. However, it now seems likely that the deviations encountered were either chance ones or the result of some artifact. Since the matching of rewards entailed a slight lag, any increase in frequency of reward appeared first in the conditional alternative and this difference could be sufficient to produce a weak preference for conditions of correlated reinforcement. The temporary preference away from a discrimination problem might simply reflect the fact that, during one stage of learning the rats had to pause to look back and forth at the discriminanda (see, e.g., Tolman, 1955), and the conditional alternative therefore took a moment longer to run. Certainly, if a powerful new principle were operating in conditional-outcome choices, its effect should have appeared more clearly in the various types of researches reported. Hence, the favored conclusion is that decisions in conditional-outcome choices can be understood on the basis of the same principles that govern nonconditional-outcome choices.

Of course, additional research is needed to establish the generality of this conclusion. Other procedures, particularly ones in which the conditional aspect is very prominent, may be more likely to detect a preference, and other organisms, particularly primates, may be more likely to reveal one. Furthermore, although a micromolar theory (Logan, 1960) implies that choice is the fundamental measure of differential response tendency, it is important to determine whether there are differences between the types of alternatives in measures such as resistance to extinction or in the effect upon them of variables such as drive. The results of these researches could deviate from the present findings but they would, at the same time, provide interesting comparative and methodological evidence.

Somewhat tentatively, then, it is concluded that there is no preference between a conditional-outcome alternative and one with the same frequency of reward uncorrelated with performance. This conclusion in turn supports the thesis that behaviors can be aggregated into a single response class even though there may be differential reinforcement among them.[2] The selection of a response depends upon the outcomes actually being encountered without regard to the fact that those outcomes may depend in some way upon how the subject performs subsequent to the choice. Accordingly, it is proposed that there are no restrictions within the conditions of reinforcement on the formation of response classes.

SUMMARY

A conditional-outcome choice is one in which the reward for at least one of the alternatives depends upon the organism's performance. A series of studies was reported in which rats were given a choice between two alleys, in one of which they had to run slowly, choose the correct brightness, or turn in the correct direction in order to obtain reward. It was concluded that rats are indifferent between such an alternative and one in which the same frequency of reward was given regardless of performance. This finding was interpreted as indicating that distinguishable behaviors can be aggregated into a single response class even though they are differentially reinforced.

REFERENCES

GUTHRIE, E. R. Association and the law of effect. *Psychol. Rev.*, 1940, **47**, 127–148.

HERRNSTEIN, R. J., & MORSE, W. H. Escape behavior maintained by the removal of a stimulus correlated with an intermittent schedule of food reinforcement. *Amer. Psychologist*, 1956, **11**, 495.

HULL, C. L. *Principles of behavior.* New York, Appleton-Century, 1943.

LOGAN, F. A. *Incentive: How the conditions of reinforcement affect the performance of rats.* New Haven: Yale Univer. Press, 1960.

LOGAN, F. A. On the specificity of discrimination learning to the original context. *Science*, 1961, in press.

SPENCE, K. W. The postulates and methods of "Behaviorism." *Psychol. Rev.*, 1948, **55**, 67–78.

TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–326.

(Received March 20, 1961)

---

[2] This thesis is not antithetical to the micromolar assumption that responses differing quantitatively are to be treated as different responses. The present position is that one can aggregate responses of different speeds into a single class for the purpose of predicting whether the rat will run down one alley or another, but the speed of running will not necessarily index the strength of that response. When different responses are aggregated into a single class, whether they are differentially reinforced or not, no statements can be made about differences among the responses within that class.

VOL. 69, NO. 6 NOVEMBER 1962

PSYCHOLOGICAL REVIEW

OBSERVATIONS ON ACTIVE TOUCH[1]

JAMES J. GIBSON

Cornell University

Active touch refers to what is ordinarily called *touching*. This ought to be distingushed from passive touch, or *being touched*. In one case the impression on the skin is brought about by the perceiver himself and in the other case by some outside agency. The difference is very important for the individual but it has not been emphasized in sensory psychology, and particularly not in the experimental literature.

Active touch is an exploratory rather than a merely receptive sense. When a person touches something with his fingers he produces the stimulation as it were. More exactly, variations in skin stimulation are caused by variation in his motor activity. What happens at his fingers depends on the movements that he makes—and also, of course, on the object that he touches. Such movements are not the ordinary kind usually thought of as responses. They do not modify the environment but only the stimuli coming from the environment. Presumably they enhance some features of the potential stimulation and reduce others. They are exploratory instead of performatory. In this respect, these touching movements of the fingers are like the movements of the eyes. In fact, active touch can be termed *tactile scanning*, by analogy with ocular scanning.

By means of active touch a great many properties of the adjacent environment can be perceived in the absence of vision. The blind must depend on it for most of their information about the world (Revesz, 1950). Nevertheless, despite its importance, the "sense of touch" (Boring, 1942, Ch. 13; Geldard, 1953, Ch. 9–12) has been studied by sensory physiologists only as a passive or receptive channel. It is treated as a part of cutaneous sensitivity. The sensitivity of the skin, which includes temperature and pain as well as touch, can be most easily investigated by the applying of stimuli to the cutaneous surface. At a perceptual level, Geldard (1957) has applied multiple stimuli to the skin and has thereby demonstrated some of the various "messages" that the skin will transmit but he is nevertheless concerned with a mosaic of receptors, not an exploratory organ. Katz (1925) and Revesz (1950) have argued, however, that the *hand* is a kind of sense organ as distinguished from the skin of the hand. Katz began the description of tactual experience as it occurs in life, and performed experiments on some of the remarkable discriminations that are possible. Revesz, observing the performances of the blind, proposed an unrecognized mode of experience called *haptics* which goes beyond the classical modalities of touch and kinesthesis. His interest is in philosophy and esthetics. These two seem to be almost the only investigators who have given much thought to what is here called active touch. Their work has not been followed up by experimenters and the term itself has never been widely used, perhaps because it does not fit with the established ideas of what constitutes a sense.

---

[1] This research was supported in part by the Office of Naval Research under Contract NONR 401 (14) with Cornell University.

The relation of touch to kinesthesis. It seems to have been assumed by psychologists that active touch is simply a blend of two modes of sensation, kinesthesis and touch proper. That is, touching merely combines the data from the feeling of movement and the feeling of contact, the two sensations being fused into one experience. This assumption, however, can be criticized. First, it fails to take account of the purposive character of touching and, second, it fails to emphasize the multiplicity of so-called kinesthesis.

The first objection is that the act of touching or feeling is a search for stimulation or, more exactly, an effort to obtain the kind of stimulation which yields a perception of what is being touched. When one explores anything with his hand, the movements of the fingers are purposive. An organ of the body is being adjusted for the registering of information. The limbs and extremities are, of course, motor organs as well as sensing organs, whereas the eyes are only sense organs, but the function of motor performance can be subordinated to the function of exploratory adjustment in the case of the limbs.

Behavior theorists have recently been emphasizing the importance of what is variously called "feedback" from responses, or self-produced stimulation, or proprioception in general. They recognize that such inputs are necessary for the purposive control or steering of behavior. What they have not recognized is that some of the stimulation produced by responses yields objective information as well as subjective. The stimulation produced by sense-organ adjustment is of this sort. In active touch, as will later be suggested, the flow of stimulation contains two components, one *exterospecific* and one *propriospecific*. But these components are not the traditional sensations of touch and kinesthesis respectively. The hypothesis will be advanced that the purpose of the exploratory movements of the hand is to isolate and enhance the component of stimulation which specifies the shape and other characteristics of the object being touched.

A second objection is that the term kinesthesis itself means different things and should not continue to imply the existence of a unitary sense. Historically the word refers to movement of the body, and originally meant the "muscle sense." But it is known that there is sensitivity to the position of the body and of all its members relative to one another. This is an articular sense, not a muscle sense, and the joints yield information about joint position as well as joint rotation (Goldscheider, 1898). Probably there is no "sense" of muscle contraction as such, the joint initiating the spatial information and the muscle spindles having only reflex coordinating functions (Rose & Mountcastle, 1959). There is sensitivity for the upright posture of the head, the linear and angular accelerations of the head, and a general "sense" of the equilibrium of the body. There is sensitivity of some kind for the force exerted by an individual, whether or not this is accompanied by hypothetical "feelings of innervation" (Boring, 1942, Ch. 14). There is also, of course, *visual* sensitivity to the changes of position of one's body in space. And there is proprioception in general, however this may be defined, as well as somaesthesia.

The single term "kinesthesis" cannot carry the burden of all the meanings that have been added to its original meaning. It fails to do justice to different inputs, to different combinations of these inputs, and to different functions of these combinations.

Types and subtypes of anatomical receptors involved in active touch. Passive touch involves only the excitation of receptors in the skin and its underlying tissue, although the patterns may be elaborate. But active touch involves the concomitant excitation of receptors in the joints and tendons along with new and changing patterns in the skin. Moreover, when the hand is feeling an object, the movement or angle of each joint from the first phalanx of each finger up to the shoulder and the backbone makes its contribution. And these inputs occur relative to a continuous input from the vestibular organs, along with the cutaneous input from contact with the ground. Presumably the feeling of an object by the hand involves the feeling of the position of the fingers, hand, arm, body, and even the head relative to gravity, all being integrated in some hierarchy of positional information.

The total flux of stimulation is enormously complex, but lawful modes of combination occur. Presumably the modes of combination of these inputs specify the difference between touching and being touched. Also certain special combinations probably carry information about the *object* being touched. In short, the so-called sense of touch as it is most often employed by using the hands involves the play of inputs coming from the whole skeleto-muscular system. Whether it is one sense or many is therefore a matter of dispute. It depends on what one wishes to call a sense. There is no single organ or structure analogous to an ear or an eye. There is no unique quality of sensation analogous to warm, cold, pain, sour, sweet, or red. Active touch does not fulfill the supposed criteria for a single sense modality. Nevertheless it provides a quite definite channel of information about the external environment. It is a type of perception that is isolable from vision, audition, taste, and smell, and it needs to be studied in its own right.

Mechanics of cutaneous stimulation, proximal and distal. The literal or proximal stimulus for touch, the immediate cause, is some sort of deformation of the skin, not necessarily a simple depression of the surface as by a tactile "stimulator," but almost any deformation of the elastic tissue. Pulling on a thread glued to the skin is just as effective a stimulus as pushing it in. The stimulus has to be a change in time, that is, a motion of the skin. Since the work of Nafe and Wagoner (1941) it has been suspected that whenever the skin becomes absolutely motionless even in some abnormal conformation, the receptors cease firing and the stimulus becomes ineffective. Outside of the laboratory this situation would be rare. A precisely constant pressure of an object relative to the skin would not occur, if only by reason of tremor or body sway.

The cause of elastic change of the skin's surface can be almost any mechanical event. Some work must be done, even if slight, but the actual mechanical happenings are so various and complex that the textbooks of mechanics do not begin to describe them. A common event is impact with a solid substance, but a puff of air or a drop of water will do as well. In hairy regions of the skin, even the slightest bending of a hair gives a stimulus. Some of the types of mechanical events that stimulate the skin, in terms of everyday physics, may be exemplified as follows.

1. Brief events: pressure, push, slap, pat, tap, prick. Note that these vary in area as well as in duration.

2. Prolonged events without displacement: vibration, stretching, kneading, pinching.

3. Prolonged events with displacement: scratching, scraping, rubbing, sliding, brushing, rolling. Note that these moving deformations vary in the degree of friction between the two surfaces, and also in the degree of lateral stretching of the skin as well as the degree of vertical depression.

These are relatively large scale events as compared to the micro-events that excite the actual end-organs imbedded in the tissue. But the latter are also mechanically complex. The ways in which mechanoreceptors are made to fire nervous impulses need not be considered here. The facts, so far as they are known, are to be found in the physiological literature.

The external causes of these mechanical events, the sources of potential cutaneous stimulation, are the objects and surfaces of the environment. They are at the distal end of a sequence that leads to the excitation of nerve cells. An objective surface of some kind and the cutaneous surface must "touch," either by movement of an object with reference to the stationary skin or by movement of the skin with reference to a stationary object.

Contact with the substratum. The "sense of touch" is not ordinarily thought to include the feeling of cutaneous contact with the earth. Nevertheless, the upward pressure of the surface of support on some part of the body provides, for every terrestrial animal, a constant background of stimulation. It is covariant with the continuous input of the statocyst receptors of the inner ear already mentioned. Together they provide what the ordinary person calls the "sense of support." The axis of gravity and the plane of the ground provide the basic frame of reference for tactual space perception. Active touch, as will appear, yields clear perceptions of environmental space in the absence of vision.

The touching of the substratum is neither active nor passive touch as defined above, i.e., neither the touching of a stationary object nor being touched by a moving object. It is instead a means of registering the stable environment with reference to which *both* the movements of one's body and the motions of objects occur.

OBSERVATIONS OF THE ACT OF TOUCHING [2]

In the simple experiments to be described, vision was always excluded. This was usually accomplished by having the seated observer put both hands under a cloth curtain at the front of a table, not by blindfolding him, or working in darkness. His manual activity could thus be observed.

When an object of any sort is placed in the hand of the observer or when his hand is placed on it, the following facts can be noted. He tends to bring the other hand to the object if allowed to do so. He tends strongly to curve his fingers around the object if permitted. He tends to move his fingers over the object in complex ways. Some of these are: tracing movements with one fingertip or several fingertips, opposition of the thumb and other fingers, and rubbing, grasping, or pressing movements of the fingers. Finally, he tends to name the object if he can do so, or to compare it with a familiar object if he cannot. Nonsense objects are almost necessary for such observations since the observer does not persist after recognizing something.

---

[2] The collaborators in this work were Susan Stanley, Lewis Greenberg, and James Caviness. The report has profited from a critical reading by Jesse Smith.

The observer seems to be trying to obtain mechanical events at the skin at various places in various combinations. The movements by which he obtains them do not ever seem to be twice the same, but they are not aimless. If the hand is conceived as a sense organ, he seems to be adjusting it. He appears to be searching for stimulus information.

So much for the manual activity. Reports of experience during active touch can also be obtained. They tend to be radically different from reports of experience during passive touch. The following facts, noted by the writer, are confirmed by other observers in this experiment, or are consistent with earlier observations.

Unity of the phenomenal object. When feeling a single object with two fingers, only one object is perceived although two separated cutaneous pressures occur. The separate "local signs" are not noticed. This holds true not only for finger-thumb opposition but also for other combinations of fingers. The unitary perception occurs when all five fingers are applied to an object, and even when two separate hands are applied to the object. In fact, 10 different digital pressures all at the same time yield a wholly unified experience. There are some limitations on this unity, as can be observed with Aristotle's illusion when a pencil is held between crossed fingers, but even then it is very hard to overcome, by the writer's introspection. This fact needs further investigation.

Stability of the phenomenal object. When sliding the skin over a corner or protuberance of an object, the displacement of the cutaneous pressure, the "tactile motion," cannot usually be noticed. The object seems to remain stationary even though the impression moves relative to the skin. Not only the object but also the whole space of the table, chair, floor, and room is felt as perfectly stable. This phenomenon has been studied in an experiment reported in the next section.

Rigidity or plasticity of the phenomenal object. When pressing a finger on a rigid surface or squeezing an object with the hand, it is difficult to notice the increase of intensity of cutaneous sensation; instead the observer is primarily aware of the substance and its resistance. Likewise, when pressing or squeezing a nonrigid object (a lump of modeling clay, a rubber ball, a handful of cloth) the observer is aware of the yieldingness, elasticity, or softness of the substance, not of the (very different) intensity of the back pressure on the skin. This fact deserves further study. Presumably the degree of force exerted is being registered by the joint-and-tendon receptors and the resulting degree of skin-pressure is being registered by the cutaneous and deep-tissue receptors, but the intensities are not experienced as such. Perhaps it is the *relation* between these different intensities that is registered. The proportion of the same retinal stimulus. Evidently the problem is analogous in tactual perception, although the solution can hardly be quite the same. When an individual actively moves his hand relative to his body, and thus relative to gravity and the surface of support, he is in contact with the earth as well as with the object, whether it be moving or motionless. A receptive system of higher order is excited during activity than the system during passivity. Hence the stimuli which have one excitatory capacity for a receptive subsystem in passive touch will have a different excitatory capacity in active touch—a different specificity. The relations between stimulation of the subsystems become significant. A lawful covariation of the inputs carries information different from that carried by isolation of the inputs, and yields different perceptions. Johannes Müller's law of the specific conscious quality of each excited nerve is evidently much too simple.

To lump one set of receptive systems together as *touch* and another as *kinesthesis,* then, is to obscure the function of the systems in combination. If both had their own unique quality of experience there would be some justification for the traditional belief in separate senses, but they do not. Anatomically different receptors may serve the same function and arouse the same experience, and, moreover, may serve different functions in different combinations. It might be better to drop the term sense and speak instead of "esthesic systems."

It should be noted that nothing has been asserted about *performatory* movements of the hand, as when an object is relocated or a tool is wielded. The perception is then subsidiary to action and the receptive systems take on still another function not here considered.

PERCEPTION OF THE TEXTURE, SUBSTANCE, AND SHAPE OF AN OBJECT BY TOUCH

Katz (1925) showed that the texture of a solid substance could be perceived by touch if there was relative motion between the surface and the skin. Ordinarily this means rubbing the fingers over the surface; perpendicular motion was very much less effective than lateral motion. A substance could be identified by its texture; the observers could distinguish a dozen kinds of paper including blotting paper, wrapping paper, and writing paper. Katz also demonstrated the perceptibility of other physical properties of a surface besides rough-smooth, including soft-hard, and various kinds of plasticity, elasticity, or viscosity.

Katz was primarily interested in the fine structure of the surface of an object, and the substance of which it was composed. He had less to say about the *shape* of a substance, which is perhaps the main characteristic of an object. Active touch, however, is an excellent channel of spatial information in that the arrangement of surfaces is readily picked up. The solid geometry of things is best got by feeling them. A number of observations illustrate this assertion.

Curvature as against planarity of a surface. Departure of a surface from the plane can be judged by an observer either by holding his hand against the surface or by running his fingers over it. The opposite qualities of convexity and concavity are immediately given. On a smaller scale, protuberances or indentations on a plane surface are identified, and this acuity is taken advantage of in Braille. Moreover a surface with positive curvature on one axis and negative curvature on the other is perceived as saddle-shaped.

The accuracy of all these judgments could be but has not been measured.

Slant of a surface with respect to gravity. When the observer's hand is placed on an adjustable palm-board outside the field of view, it is found that he can judge its slant with precision. He can nullify this slant and adjust it to a normal plane. This holds for all three planes of space, the frontal, the saggital, and the horizontal. Variable and constant errors for settings to the normal plane seldom exceed 2°–3°. Furthermore, the slant of a surface that is visible only can be equated to the slant of a surface that is tangible only. The cross-modal matching of slant is accepted as natural by all observers, and the manual judgment can be used as a criterion of visual perception (Gibson, 1950).

When the palm-board is maintained at a slant from the frontal upright plane for some time, "normalization" seems to occur, and a negative aftereffect ensues. In this respect, the tangible slant of a surface behaves like the visible slant of a surface (Bergman & Gibson, 1959).

Parallelity of two surfaces. If a pair of adjustable palm-boards is held between the two hands, behind a curtain, they can be adjusted to the parallel position with small errors. The divergence or convergence of the two planes are felt as opposite qualities of a simple spatial experience. This dimension is also susceptible to normalization; after a few minutes of feeling divergence (say) the setting accepted as parallel is found to have shifted several angular degrees toward divergence. This fact has been studied in detail (Gibson & Backlund, 1963).

Distance between two surfaces. The span of the hands holding an object between them is experienced vividly. The tangible extent can be matched with a visible extent or vice versa. Likewise the distance between the thumb and one or more of the opposable fingers is perceived, and is accurately matched with the width of a visible object. According to Kelvin (1954) the cross-modal size comparison is made with no less accuracy than either unimodal size comparison alone. The finger-span of one hand can be immediately felt as equal or unequal to the finger-span of the other hand, and this comparison has been utilized by Köhler and Dinnerstein (1949) in studying a phenomenon they termed the kinesthetic figural aftereffect. In the many repetitions of this experiment the equivalence of the two hands for perceiving a spatial extent seems to have been taken for granted without its being studied in its own right.

Plane-angle, or edge of a surface. To exploratory touch, a junction of two planes constituting an edge is very noticeable. The angle is perceptible as varying from sharp to dull. At the extreme case of a knife-edge, sharpness is more accurately judged by touch than by vision.

Edge-angle, or corner of a surface. Where two edges intersect, or where three or more planes adjoin, a corner is produced. In the case when the angles of a convex corner are acute it is apt to be called a point. Corners and edges seem to be the main identifying feature of many objects. The fingers appear to seek them out and explore them.

Discussion. Planarity, curvature, slant, parallelity, span, edge, and corner may be conceived as variables of solid geometry. They are all perceptible to active touch, and appear to be simple qualities of experience. Since they are amenable to physical measurement, psychophysical experiments would be possible, but few have been performed.

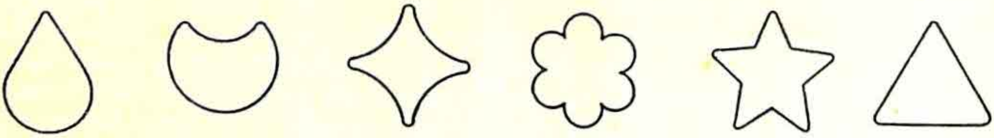


FIG. 1. The six forms to be identified by touch.

Tactual Form Perception: A Comparison of Active and Passive Touch

The study of the perception of objects by touch has been largely confined to investigations of two dimensional form or pattern on the skin, with passive stimulation (e.g., Zigler & Barrett, 1927). They are in the tradition of experiments on cutaneous localization and the two point threshold, the assumption being that the skin is first of all a receptor mosaic with patterns of excitation being projected to the brain. The forms pressed into the skin were called *solid* or *outline* depending on whether the edges of a surface or the edges of a metal strip were used; the latter gave better discrimination. The identification of forms in this manner is not very accurate, and there is some uncertainty even about circles, squares, and triangles.

Everyone knows, however, that a better way to perceive a thing of this sort is to run one's fingers over the edges. The fact is often taught to beginning students, and is even demonstrated in laboratory exercises (Langfeld & Allport, 1916) but a search of the experimental literature failed to locate its source. Lashley (1951), for example, asserts as if it were common knowledge that the same shape which can scarcely be distinguished with passive touch "can readily be distinguished when explored by tactile scanning" (p. 128). An experiment was performed to quantify this assertion, since no authority for it could be found. Each of a set of tactual forms had to be matched to its visual equivalent. The accuracy with which this could be done was compared when the form was pressed into the palm of the hand and when it was held above the palm and explored by the fingertips.

Method

The sources of stimulation were bent metal strips with handles, the edges having the forms shown in Figure 1. They were in fact small "cookie cutters" with a mean diameter of 1 inch. The six used were chosen for being mutually different in about the same degree. No two had the same number of corners or points. Each subject was shown the set of stimulators and the equivalent set of numbered drawings hung on the curtain in front of him. He then put his hand palm up on the table behind the curtain and was either touched by the object or was allowed to touch it for several seconds. In the one condition, he was instructed to open and relax his hand; in the other to cup the fingers upward and prepare to feel the edges. He then had to identify the form by number. The stimulator was not necessarily presented in the same orientation as the drawing.

The six forms were presented five times each (but the subject was told not to expect equal frequencies) and under two conditions, making a total of 60 trials. The order was random. No preliminary practice or knowledge of results was given. Twenty subjects were tested.

Results. A chance level of judgments would be 1/6 or 16.7%. For passive touch the mean frequency of correct matches was 49%. For active touch the mean frequency was 95%. The difference is significant.

In the active condition, the subject was permitted to explore the edges in any manner he chose. One style was to use all five fingertips simultaneously, another was to move the tip of the index finger around the edge, and others were intermediate. All subjects, however, actively moved their fingers and thus obtained relative motions between the skin and the object.

In the passive condition, the form was pressed down on the palm by manual action of the experimenters so that a certain amount of unsteadiness was inevitable. When subsequently, with 20 different subjects, a mechanical lever system was used to apply the stimulation the mean frequency of correct matches fell to 29%.

Some subjects reported that they tried to count the number of corners or points they could feel, and that this was easier with the active than with the passive condition.

Discussion. A common belief has been verified but, in a sense, no claim can be made for the conclusiveness of the comparison. It could be argued that the fingertips are more "sensitive" than the palm of the hand, meaning that localization is better and the two-point threshold lower. This is true. The comparison is not very meaningful. The real point of the experiment is that tactual form perception in one meaning of that term does not depend on the pattern of local signs on the skin. If any number of fingertips from one to five can be used indifferently, the chaotic pattern of local signs resulting can hardly be the sole basis of the perception. With active touch no forms existed on the skin, but only a changing pattern of pressures. Evidently the registering of "form" by touch is not as straightforward a process as experimenters have supposed. Another experiment may help to clarify the matter.

TRANSPOSABILITY OF FORM ON THE SKIN

Gestalt theorists have emphasized the fact that a form can be moved over the retina, completely changing the set of receptors stimulated, without changing the perception. The contribution to perception of the anatomical pattern of excitation becomes a puzzle. The same is true of the skin.

Along with the experiment just described, another was carried out at the same time in order to compare passive static form perception with passive moving form perception. In these trials the "cookie cutter," instead of simply being pressed into the palm of the hand, was pressed and continually rotated clockwise and counterclockwise with a twisting motion. The same area of skin was stimulated in both cases and the same anatomical receptors were available, but in the latter case the form of stimulation changed from one instant to the next.

Judgments of the form of the object were more accurate when it was rotated than when it was not, the mean frequencies of correct matches being 72% and 49%, respectively. The difference is significant at the 2% level. To the writer's introspection, the corners of the object seemed to "stand out" better when it was twisted. In short, the form of the object seemed to become clear when the form of the skin-deformation was most unclear.

Discussion. Continuous change of the proximal stimulus with no change of the rigid object constituting the source of the stimulus is an old problem in visual perception. Why does the perception correspond to the form of the object instead of to the form of the stimulus? Evidently a quite similar problem arises for tactual perception. The paradox is even more striking, for tactual perception corresponds well to the form of the object when the stimulus is almost formless, and less well when the stimulus is a stable representation of the form of the object. A clear unchanging perception arises when the flow of sense impressions changes most.

It might be that the skin does not have as its primary function the registering of form as this has usually been conceived. Impressions of form and location relative to the skin might be quite separate from and incidental to the use of the extremities as sense organs. The informative stimuli might well be incorporated in the seemingly complex motions and transformations of the skin. There must exist relations of separation and proportion which remain invariant over time, and which are specific to the object. They do not represent the shape of the object but they specify it. The solution to the problem of object perception in touch would then be that continuous change in the proximal stimulation is accompanied by nonchange, that is, the set of invariant relations. The former is not noticed; the latter is separated out and attended to.

The role of exploratory finger movements in active touch would then be to isolate the invariants, that is, to discover the exterospecific component in the flux of stimulation. Only thus could the above paradox be resolved.

RELATION BETWEEN TOUCH AND VISION

A number of similarities and differences have been noted in the course of these observations between active touch and visual perception. Many of these have been pointed out before by other observers such as Katz (1925, 1935) but a few of them have not previously been emphasized.

Surface perception. Solid surfaces are tangible and visible. This can only be because they have mechanical inertia in the first place, and because they reflect light with scattering in the second place. (A few visible surfaces such as smoke are not tangible and a few tangible surfaces such as plate glass may not be visible, but most surfaces are both.) The physical texture of a surface, the various kinds of roughness, can be observed either by vision or by touch. The color of a surface is visible only, not tangible, and consequently the pigment-texture of a surface (the differential coloring of marble or of a picture) is the same. On the other hand, if color is intangible temperature is invisible. Each sense has its special sensitivities to the properties of a surface but they also have some in common. The bark of a tree looks rough without touching it and feels rough without seeing it. The expert can identify the tree by either sense alone.

The visible array of surfaces extends for miles of course, and the tangible array only to the length of man's arm. But if surfaces determine space perception, both touch and vision are spatial senses.

Object perception. A detached solid substance with a topologically closed surface is perceptible as such to both the hands and the eyes. By virtue of physics, the eyes can encompass the closed contour of a very large object (depending on its distance) while the hands can explore only the surface of an object of limited span. The unity of the visual perception is said to be based on the "figure-ground" phenomenon, the simultaneous registering of the whole contour, whereas the unity of the tactual perception has to be based on either cutaneously separated impressions or on successive impressions. Vision is said to register only the front surface of an object while touch registers the front and the back at the same time. In these respects vision and touch seem to differ. Nevertheless these contrasts are exaggerated; succession enters into the operation of both senses. The eyes normally fixate in succession just as the fingers explore in succession. With two eyes, and by changing one's standpoint, more of an object than its front surface is perceived. As regards the unity of separated impressions from the two hands, there is a visual analogy in the unity of separated impressions from the two retinas.

Spatial properties of objects. Certain variables of the shape of a detached object, slant, curvature, edge and corner are independently visible and tangible. The equivalence of the two modes of perception for judgments of the object is such that differences got by one sense are equated to differences got by the other. The stimulus information for these cross-modal judgments, one depending on mechanical and the other on luminous energy, might seem hopelessly discrepant, but the fact is that the judgments can be made.

Whether both vision and touch can register form is an ambiguous question, as the experiments showed. The answer depends on what is meant by form. Drawings and pictures on a flat surface are sensed only by vision. The conformation of the edges of an object, however, is sensed by both vision and active touch. The tangible and visible features of things abstracted by solid geometry (slant, curvature, edge, corner) did not seem any more complex to introspection than the solely visible features of things abstracted by plane geometry (triangle, square, circle).

A new set of shapes has now been devised for the study of object-perception by active touch. They consist of ten solid sculptures, or free-forms, made of plastic, the surfaces being curved, with no planes, edges, or corners. They are intended to be felt with two hands (and are called "feelies"). Approximately one-half of each surface (the "rear") is convex; the other half (the "front") consists of six convexities with intermediate saddles or concavities. In general, there are five protuberances around a central protuberance, but no object is symmetrical, either radially or bilaterally. They cannot, therefore, be distinguished from one another by counting. Each is readily discriminated from every other by vision of the "front" surface. They are also mutually distinguishable by feeling, although with some error and hesitation for an unpracticed observer.

Replicas of the ten objects are available, made from the same molds. It is therefore possible to present one object to the hands and the same or a different object to the eyes simultaneously. Preliminary results indicate that the cross-modal matching of these novel preceptions is possible even for a naive observer, and that practice can bring about errorless judgments in all observers so far tested.

Passivity and activity in touch and vision. In passive touch the individual makes no voluntary movements. Similarly, in passive vision he makes no eye movements, which means that he must voluntarily fixate his eyes on a point specified by the experimenter. Neither state is natural to an individual. In a tactual situation, the observer will explore with his fingers unless prevented and, in a visual situation, he will explore the focussable light, fixating, accommodating, converging and pursuing. Both senses are normally active. The passive stimulation of the skin or the retina is necessary for the study of the receptor-cells in the skin or the retina, but the experiences resulting are atypical.

In active touching and looking the observer reports experiences of a quite different order. They correspond to the environment instead of to the events at the sensory surface. The experiences noted with passive stimulation can scarcely be noticed, if at all. The effective stimuli are presumably different, being relations and combinations selected from a flux of potential stimuli. Moreover, the experiences of active touching and looking are more nearly equivalent; they are more alike to introspection than those of passive touch and vision.

In general, the foregoing survey suggests that vision and touch have nothing in common *only when they are conceived as channels for pure and meaningless sensory data.* When they are conceived instead as channels for information-pickup, having active and exploratory sense organs, they have much in common. In some respects they seem to register the *same* information and to yield the *same* phenomenal experiences.

CONCLUSIONS AND SUMMARY

A series of observations, both introspective and behavioral, confirms the distinction between *touching* and *being touched*. The former is a channel for a great variety of information about the environment, but whether it should be considered one or several senses is a matter of definition. The simple formula that it consists of passive touch plus kinesthesis is insufficient. The hypothesis of two components of stimulation, one exterospecific and one propriospecific, is more promising.

Many properties of surfaces and objects can be registered by active touch. The accuracy of such judgments can be measured and formal experiments are needed. The observations suggest that the variables of solid geometry can be more profitably studied than those of form. Object constancy is characteristic of tactual perception. In several respects, tactual and visual perception are alike.

In general, experimenters have not realized that to *apply* a stimulus to an observer is not the same as for an observer to *obtain* a stimulus. Obtainable stimuli can be controlled and systematically varied, just as applied stimuli can be. Psychophysical experiments are possible for tactual perception as well as for cutaneous sensitivity if appropriate methods are devised for controlling the obtainable stimuli.

REFERENCES

BERGMAN, R., & GIBSON, J. J. The negative after-effect of the perception of a surface slanted in the third dimension. *Amer. J. Psychol.*, 1959, **72**, 364–374.

BORING, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century, 1942.

GELDARD, F. A. *The human senses.* New York: Wiley, 1953.

GELDARD, F. A. Adventures in tactile literacy. *Amer. Psychologist*, 1957, **12**, 115–124.

GIBSON, J. J. The perception of visual surfaces. *Amer. J. Psychol.*, 1950, **43**, 367–384.

GIBSON, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.*, 1954, **61**, 304–314.

GIBSON, J. J., & BACKLUND, F. An aftereffect in haptic space perception. *Quart. J. exp. Psychol.*, 1963 (in press).

GOLDSCHEIDER, A. *Gesammelte Abhandlungen von A. Goldscheider.* Vol. II. *Ueber die Bewegungsempfingung.* Leipzig: Barth, 1898.

JAMES, W. *Principles of psychology.* New York: Holt, 1890.

KATZ, D. *Der Aufbau der Tastwelt.* Leipzig: Barth, 1925.

KATZ, D. *The world of colour.* (Trans. by R. B. MacLeod & C. W. Fox) London: Kegan Paul, 1935.

KELVIN, R. P. Discrimination of size by sight and touch. *Quart. J. exp. Psychol.*, 1954, **6**, 23–34.

KÖHLER, W., & DINNERSTEIN, D. Figural after-effects in kinaesthesis. In, *Miscellanea psychologica Albert Michotte.* Louvain: Editions de l'Institut Superieur de Philosophie, 1949. Pp. 196–220.

LANGFELD, H. S., & ALLPORT, F. H. *An elementary laboratory course in psychology.* Boston: Houghton Mifflin, 1916.

LASHLEY, K. S. The problem of serial behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951, Pp. 112–136.

NAFE, J. P., & WAGONER, K. S. The nature of pressure adaptation. *J. gen. Psychol.*, 1941, **25**, 323–351.

REVESZ, G. *Psychology and art of the blind.* New York: Longmans Green, 1950.

ROSE, J. E., & MOUNTCASTLE, V. B. Touch and kinaesthesis. In *Handbook of physiology.* Washington, D. C.: American Physiological Society, 1959. Ch. 27.

ZIGLER, M. J., & BARRETT, R. A further contribution to the tactual perception of form. *J. exp. Psychol.*, 1927, **10**, 184–192.

(Received July 30, 1961)

Psychological Review
1962, Vol. 69, No. 6, 492–500

A SYSTEM FOR SHAPE RECOGNITION

J. A. DEUTSCH[1]

Department of Psychiatry and Psychology, Stanford University

In this paper it is suggested how certain morphological features of the histology of the visual system may on simple assumptions function to discriminate distance, angle, and shape. For the purposes of this paper shapes can be regarded as a derivative of distance between points on the contours and also as a derivative of the angles between parts of the contour. It is difficult to know whether in the human case it is the relative distance cues or the cues from angles which are more relied upon. In the case of a visually more primitive organism, such as the rat, though it is known that it can discriminate distance and also relative size, its abilities to discriminate the angle of one line to another or the curvature of lines remains untested. It is therefore entirely possible that such animals lack a mechanism to enable them to discriminate angles. If this is the case the shape discriminatory abilities of such organisms are based mainly on comparisons of relative distance between contours.

It would seem likely that in discriminating shape animals utilize prior information available in the nervous system about distances between contours and about angles between contours also. However, the basis of the capacity to estimate distance between any two points in the visual field or the angle between any two contours in the visual field is obscure. If we wish mechanically to assess the relation between two things or events such as the distance between two points, it is necessary (though not sufficient) that these two things or events jointly determine some third event.

On the face of it, a system of fixed connections would appear to be very uneconomical in the number of structural elements which would have to be employed to assess distance (Sutherland, 1959). Suggestions have therefore been made to decrease the number of elements required to relate visual points or contours so that they can jointly determine some third event. One such way is to transform distance in space into distance in time by some kind of moving process, a notion familiar from television. One such suggestion of scanning from fixed coordinates external to the shape to be analyzed has come from Dodwell (1957) for the rat and Deutsch (1960) for the octopus. Deutsch (1955, 1958, 1960b) has suggested that such a conversion of distance into time is effected not by a scan external to the shape to be analyzed but by a wave of excitation propagated by the contours of the shape itself.

However, further investigation of the problem has led the writer to the conclusion that a system relying on a conversion of spatial separation into temporal interval is not necessary. For instance an examination of the neurohistological data concerning the zona plexiforme of the deep retina in the octopus has led him to postulate a system for shape recognition which does not rely on these principles (Deutsch, 1960a), though no functional fixity of connections has been postulated. There is good correspondence between this system and what can be seen of the striking morphology of the zona plexiforme and the properties of the shape discriminating system of the octopus as they have been ascertained so far. However, the shape discriminatory capacities of the octopus are different from those found among the vertebrates. It seems nevertheless that even in other cases besides the octopus it is possible to devise a system of reasonable economy in which the principle of transforming distance into time is not necessary.

---

[1] The writer is indebted to John Gilbert for help in computing the "outputs" of the suggested system. He also wishes to thank G. A. Horridge and T. H. Bullock for valuable discussions.

SPECIFICATIONS FOR A BIOLOGICAL SHAPE RECOGNIZING MECHANISM

In a previous paper on the subject (Deutsch, 1955) the present author enumerated some of the peculiarities of shape recognition in animals which a theory should explain. These peculiarities are not common to all animals, such as the octopus, and here special theories are necessary (such as Deutsch, 1960), but these peculiarities appear to be those of the general case of higher vertebrates. (Unfortunately the experimental evidence in this field leaves much to be desired. But if it is argued that the properties set down below have not been established experimentally to be the case, then we can regard the properties listed below as predictions instead from the theory from which they are deducible.)

These are some of the abilities that a system must have before it can be considered a theory of shape recognition.

1. It must be able to produce the same output independent of the location of the image on the receptor surface. Shapes retain their identity in spite of changes in their position in the visual field.

2. The system must produce the same output independent of the rotation of the shape on the receptor surface. (For some of the evidence on this point see Gellermann, 1933, and Deutsch, 1960b).

3. A similar output should be given for the same geometrical shape of different sizes, but in such a way that transfer to a larger shape is more efficient than transfer to a smaller shape.

4. Mirror images should produce an identical output. They cause great difficulty in discrimination in animals and children and are readily confused even by adults.

It should at this stage be emphasized that only a partial set of visual classifying mechanisms is being suggested. The fact that we can perform visual discrimination of the tilt, position, or brightness does not constitute a disproof of a theory of distance, angle, and shape recognition. It is clear that the mechanism which classifies such properties per se must be discarding information about tilt, position, or size. It must be an abstracting mechanism par excellence. The fact that an animal can visually discriminate other properties besides these simply means that it is also possessed of other visual mechanisms. Therefore, if an animal can be shown to discriminate between two stimuli this does not disprove the notion that a particular system is not to be found inside the animal; it could be supplemented by another mechanism. On the other hand an inability to discriminate, which would be predicted by the theory, would tend to support it as would a tendency to transfer to other stimuli, identical on the theory, even where discrimination can be taught.

Let us begin by examining a neurohistological picture (Figure 1) of the bee's optic lobes (from Cajal & Sanchez, 1915) which gives a structure

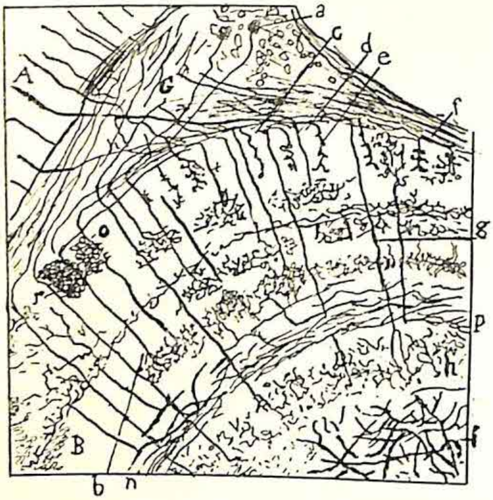


FIG. 1. Neurohistological picture of the bee's optic lobes.

well suited for the transformations to be postulated.

> As all the authors have recognised, the plexiform mass of this lobe appears to be concentrically furrowed, by means of concentric bands, of a finely granulated aspect, which appears to Radl to constitute the anatomical characteristic of the visual centres of the animal series, and which remind not a little of the plexiform layers of the optic lobes of the fishes and batrachians, so minutely described by P. Ramon (pp. 99–100, author's translation).

Such a structure is not unique to the bee, though the particular section here is unusually good for the purpose of illustration. It is to be observed that the lobe is divided by various horizontal layers, three of which are indicated (at g, p, and i). The retina is not in the section, but fibers from it are shown proceeding from the direction of A (upper left-hand corner) which enter the lobe at right angles to the horizontal layers after crossing at the chiasma. It can be seen that in the first layer to be reached by the retinal fibers the number of terminations is large, and the size of the terminal arborizations is small by comparison with those further inside the lobe. From this it is possible to infer that in this layer is preserved information about the activity of a larger number of ommatidia than in layers further in and also that the ommatidia which are represented in this layer are closer neighbors to each other spatially in the eye than those represented by fibers which terminate on subsequent layers. The fibers proceeding from the retina entering further layers are fewer, presumably from points further apart on the retina, and their terminal arborizations are larger than in the previous layers, thus giving scope for interaction over a wider angle. So it seems that as we go inwards from the retina, progressively fewer fibers cross each horizontal layer and even though the fibers are converging (because of the concentric arrangement of the layers) the extent of the terminal arborization at each level progressively increases.

Given such an anatomical arrangement, we need to make some assumptions about how the parts behave in order to speculate about the functions of this arrangement. The only assumption which is necessary is that the fibers in each horizontal layer act as spatial summators. We need only postulate that if two incoming retinal fibers which are spatially contiguous are also simultaneously excited, then this will excite fibers in the horizontal layer. The sum of the excitation generated in this manner in each layer will reflect the number of neighboring fibers that are simultaneously active.

As the fibers which are neighbors in the various layers are drawn from angular distances progressively further apart, there will be distinctive patterns of activity emerging from the horizontal layers depending on the overall distribution of widths and breadths of visual areas. A totally uniform excitation of all the visual elements will cause maximum activity in all the horizontal layers. A visual field composed of contrasting areas each of which spans that angle of field from which one receptor is selected to end in a particular horizontal layer will produce a minimum of activity in this layer. In such a case, alternate fibers ending in this layer will be inactive, and so spatial summation will not occur. As we inspect the output from horizontal layers which sample closer neighborhoods, the rate of activity in these will increase as we move away from the layer in which activity is at a minimum. The same will occur as we move in the other direction where fibers drawn from more distant points end in horizontal layers further down.

The system so far described will perform a kind of crude autocorrellogram of an area with itself, indicating the number of times areas of a given size repeat regardless of direction. Each horizontal layer performs this task for an area of one approximate size. Therefore, a uniform excitation of the whole visual field should produce a maximum of excitation on all layers. On the other hand, a checkerboard with squares of a certain size should cause great lowering of excitation in one particular horizontal layer, as the repetition frequency (taking all directions together) of an excited area will of necessity be small. It can be seen that the system described for the bee could not perform the type of discriminations to be seen in the rat, for instance. However, we have supposed, in order to be economical in our assumptions, that it samples receptors from gradually increasing distances at each layer. It is evident that such a system would perform very poorly when confronted with outline shapes and best when dealing with regularly repeated blobs or stripes. This property tallies with what is known of the bee's visual abilities. It is not supposed that the vertebrate system is of precisely the same type. The example of the bee is quoted in order to show that the kind of structure necessary to perform our postulated transformation is a biological possibility. This structure can without much difficulty be assigned the kind of task which we require.

The higher vertebrates appear to take contours as the basis of their shape discrimination. Therefore, neighboring arborizations of fibers at successive layers should be made to send a pulse to the horizontal layers only when both of them are active and a fiber should only be made active if it has been stimulated by a contour discontinuity. Further, fibers descending to progressively lower layers cannot merely be samples of retinal stimulation and progressively larger distances, but must give information about the presence of a contour at within these distances. Given these alterations, we have a system which is capable of discriminating distances. A particular distance between two retinal points will be represented as activity in a certain horizontal layer. As a result it can be used as the basis for a certain kind of system of shape discrimination.

System. The basic function the postulated system performs is to measure the distance between all pairs of points on the boundary of the shape and so produce a distribution of how many pairs of points there are at each distance from each other. This may be made clearer by an example. If we take four points at the corners of a square, we can then measure the distances between all the pairs of points. The histogram representing this figure will show a total of four at the distance equal to one of the sides of the square and a total of two at the distance equal to the length of the diagonals. We can perform a similar transformation of a

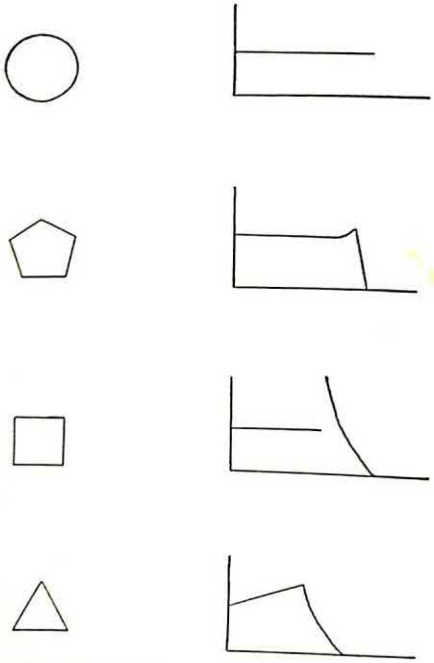


Fig. 2. Histogram measuring the distance between all the pairs of points.

shape given any number of points on its boundary to derive the curves shown in the illustration (Figure 2). As can be seen, different shapes under such a transformation still remain different from each other.

Given that any two points of a certain distance from each other will cause activity on one particular horizontal layer, then if there is more than one pair of points at that distance from each other, more activity on that one particular horizontal layer will result. Given a set of points at varying distances from each other, activity on successive horizontal layers will show the distribution of these distances. By inspecting the output of these layers we have the information of how many points there are at the distance from each other, at the distance +1, and so on. We thus have a system performing the type of transformation postulated above and which also conforms to all the specifications (except Number 3) laid down above. It should be clear that the first specification has been obeyed. The same output will be produced by the horizontal fibers independent of the position of the vertical fibers impinging on the horizontal layer with their arborizations, given that channels equal distances apart retinally end on the same horizontal layer. Similarly, rotation of the optical input will not produce any change unless we postulate special and regular distortions of the distribution of the receptor surface or special collecting properties of the horizontal layers. Finally, mirror images will produce the same output. The output is simply a measure of the distances between the points in a figure and these are the same in mirror images.

The specification which has not been fulfilled is that relating to the invariance of the same shape in spite of different size. This requires some further considerations. We have seen that in the optic lobe of the bee that the number of fibers relative to the distance along the retina decreases at each successive layer as the fibers travel from the retina. In this way the output from the horizontal layers which would be sending signals out in succession as two lines of fixed length were moved farther and farther apart would decrease layer by layer. Now the successive thinning out of the fibers could be such that the firing from these layers would not decrease as the lines were increasing in length in proportion to their distance. If we descend layer by layer (measuring the number of neighbors at progressively larger distances) the output generated by a small shape will be the same as the last portion of the output generated by the same larger shape. In this way the pattern for a small shape will be completely reproduced if we transfer the animal to a discrimination to larger shapes, though it will be added to at the beginning, whereas if we transfer to smaller shapes, the larger shape to which the animal was originally trained will have the first part missing. The problem of how the same relative message appearing on different horizontal layers is then taken and reacted to as the same message is still of course unsolved. There are various alternatives, but it is probably better to leave this problem by saying that the third specification looks as if it could be fulfilled. Shape would be given by the relative position of the signals on the horizontal layers to each other, size by their absolute position on these layers.

A remark which should be made about the system as it stands is that any coincidences (signaling that two contour components are within a certain distance) will be repeated at each succeeding horizontal layer, though they will not be signaled at levels which measure smaller distances. There are ways of removing this redundant information. For instance, we can alter our system in such a way that all the fibers from the retina are taken through all the layers with increasing overlap of arborization at each level. We can then set the system in such a way that only the boundary or end of an arborization and the center of another arborization (that is another fiber) will produce a response in the horizontal layer when they are both on. Though there will not be a reduction of the number of fibers traveling down to each succeeding level in this case, we may still suppose that the area over which fibers interact to cause a signal in the horizontal layer increases as we descend downwards.

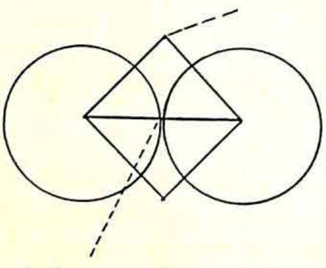
ANOTHER MECHANISM

The mechanism which was derived from the neurohistological picture of the bee's optic lobes is not the only kind of system which would produce the postulated transformation and satisfy the stated specifications. The suggestion may be worth making as neurohistologically distinct variants of the same principle could have developed.

This second mechanism would also generate the same signals, given the same shape, irrespective of its locus on the receptor surface and its tilt on the frontal parallel plane. Further, it will confuse mirror images in any orientation (as a young child will). Also, this mechanism will reduce shapes of different size to a similar message in the same way as the mechanism described above. That is, the message for a small shape will be the same as one end of the message signaling the same but larger shape. This would be consistent with the well known asymmetry of transfer between smaller and larger shapes. It is easier to obtain transfer to a larger shape from a smaller than vice versa. When transferring to a larger shape the animal will obtain the complete message it has learned together with an additional portion. However, when transferring to a smaller shape, some of the message representing the shape it has learned will be missing. We might therefore reasonably assume that transfer would be worse in this second condition.

In the mechanism to be considered, let us assume that there is a three-dimensional space filled with cells. Each cell on the top of this solid sends down connections into the solid. Each cell or point is conceived as radiating a large number of connections in all directions, but still at a constant angle. Each point at the top may thus be conceived as the apex of a cone, and the connections as lines on the surface of this cone converging at the apex of the top of the solid. (We can, for convenience, think of these as lines regularly spaced on the surface of the cone.) When a contour is projected on the top of this multicellular solid, each cell at the base will send down excitation into its set of connections fanning-out downwards as on the surface of a cone. If we picture such ascending excitation as black lines on the surface of each cone, these lines on two different cones will intersect at a height depending on the distance of the apices of these cylinders. To place a limitation on the number of intersections of lines between the two cylinders to two, we shall arbitrarily assume that only head-on collisions between lines of excitation will be registered as coincidences. That is, only lines at a 180-degree angle to each other will be counted as intersecting or, in other words, lines intersecting at the tangent of the circumference of the two cones.

this not to be counted at a higher level



only this to be counted as coincidence

In this way the distance of two points at the top of the solid will determine the height at which the excitations emanating from each will coincide. The number of coincidences at a particular height or horizontal layer will be determined by the number of points (or rather cells) at a particular distance from each other on the contour of a particular shape. Thus the arrangement outlined above will produce a histogram or a spectrum for each particular shape, with distances along one ordinate and the number of points at each distance from each other on the other ordinate. The spectrum or histogram produced will be the same independent of the position of the shape at the base of the solid and also independent of its orientation. Mirror images will also give rise to the same spectrum. The property of generalization to the same shape of different sizes is yet to be accounted for. Here it is possible to have recourse to the same kind of principle as in the first type of mechanism outlined above. We need merely suppose that the number of possible coincidences decreases by the factor of the distance downwards. For instance, there may be a certain critical distance between two ascending lines as they pass near each other within which a coincidence signal will be generated if both lines are excited. As the lines pass downwards on the cone they will be progressively more widely spaced on the surface of the cone and therefore increasingly less likely to be within the critical distance of each other, even if their angle with regard to each other is sufficiently correct. If, furthermore, the coincidences are counted by layers progressively further apart (in terms of retinal distance) the histogram generated by a small shape will be the same as the end of the time series of the same large shape.

The advantage of this second scheme is that it could be made to yield good discriminations between various degrees of curvature. This would occur if we assume that once coincidence had occurred between two paths of descending excitation, the excitation would not descend any further. In this way the excitation generated by a straight line would all coincide on the first layer and there would be no further coincidences whereas an increase in curvature would bring about further coincidences (Figure 3).

Discrimination of Angle

So far two systems have been described which yield discrimination of distance between points. Points are brought into relation at some third point and this third point varies along a single continuum depending on the distance between the two points. Further, all the pairs of points with the same distance between them are represented by activity in the same structure. This structure is conceived of as a layer, so that a particular distance between any two points is translated into activity in a particular layer. It was shown that shape recognition could be based on such a system. However, as was stated above, discrimination of size of angle between two contours is also a capacity of the visual nervous system. It will be shown briefly below how the type of mechanism already outlined can simply be expanded to provide this capacity.

It will be apparent that the rate at which two contours converge or diverge will be reflected in the angle at which successive coincidences between points occur in the horizontal layers. For instance, two parallel lines will cause a set of coincidences which is horizontal to the successive layers. An increase in angle between lines will cause a gradual tilting away from the horizontal of the successive points of coincidence. Such information could be read out from the system in different ways. One such possibility would be to have cells with fields of arborizations shaped like discs or flattened spheres sending a signal out whenever two coincidences occurred within the field. With considerable overlap of these fields, angle would be measured by the number of cells sending a signal out. If this information was sent out layer by layer, a mechanism analogous to the one based on distance could operate on the basis of angle. Such functioning of cells and their arborizations is not implausible in view of the work of Maturana, Lettvin, McCulloch, and Pitts (1960) on the frog's visual system.

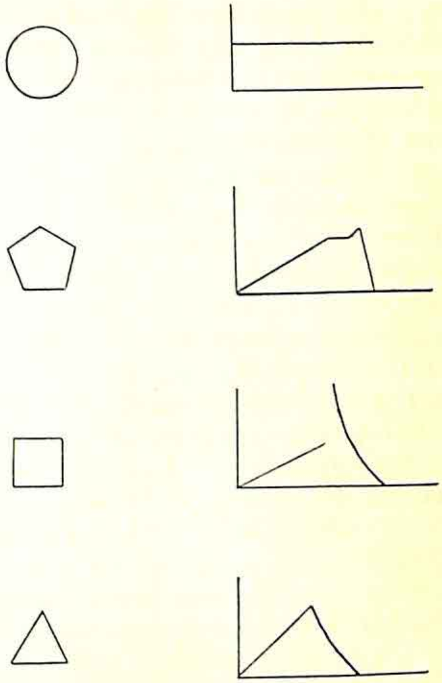


FIG. 3. Histograms produced by second mechanism.

DISCUSSION

The question of uniqueness of transformation of a given shape has not been so far dealt with. No mathematical proof is so far forthcoming that shapes or patterns other than left-right or up-down mirror images are uniquely transformed. However, extensive tests of a trial-and-error nature have failed to reveal different shapes with identical profiles, though shapes that appear similar give rise to profiles with certain common features.

The "preset" systems described above give rise to certain determinate principles of classification, and so of similarity. Such principles seem to operate in the case of our own shape recognition and therefore they were set up as specifications for a theoretical system. The similarities were therefore built into the system, in contrast with such attempts as those of Uttley (1956) and Taylor (1956). Here equivalences must be learned. Unless the machine has some already built in classification systems to enable it to abstract the principles on which it is being made to equate diverse groups of dots, its economy is bound to be low as compared with a preset system, which already abstracts the same properties as the human operator. Its performance in novel situations is also likely to be poor. Sutherland (1959) has made a strong case for supposing that there is a preset classifying mechanism. If there is, it is important from the point of view of mechanical shape recognition to find out what it is. The ideal machine would classify shapes in the way we do and so would see similarity in distorted human output, such as handwriting, where distortion is permitted because it seems that the essential elements of identity are not lost.

REFERENCES

CAJAL, S. R., & SANCHEZ, D. *Contribucion al conocimiento de los centros nervosos de los insectos.* Vol. 13. Madrid: Universitad de Madrid, Trabajos del Laboratoria de Investigaciones, 1915.

DEUTSCH, J. A. A theory of shape recognition. *Brit. J. Psychol.*, 1955, **46**, 30–37.

DEUTSCH, J. A. Shape recognition: A reply to Dodwell. *Brit. J. Psychol.*, 1958, **49**, 70–71.

DEUTSCH, J. A. The plexiform zone and shape recognition in the octopus. *Nature, London*, 1960, **185**, 443–446. (a)

DEUTSCH, J. A. *The structural basis of behavior.* Univer. Chicago Press, 1960. (b)

DODWELL, P. C. Shape recognition in rats. *Brit. J. Psychol.*, 1957, **48**, 221–229.

GELLERMANN, L. W. Form discrimination in chimpanzees and two-year-old children. I. Form (triangularity) *per se.* *J. genet. Psychol.*, 1935, **42**, 3–27.

MATURANA, H. R., LETTVIN, J. Y., McCULLOCH, W. S., & PITTS, W. H. Anatomy and physiology of vision in the frog (*Rana pipiens*). *J. gen. Physiol.*, 1960, **43**, 129–176.

SUTHERLAND, N. S. Stimulus analysing mechanisms. In, *Symposium on the mechanization of thought processes.* London: Her Majesty's Stationery Office, 1959.

TAYLOR, W. K. Electrical simulation of some nervous system functional activities. In C. Cherry (Ed.), *Third London Symposium on information theory.* London: Butterworth, 1956. Pp. 314–327.

UTTLEY, A. M. Temporal and spatial patterns in a conditioned probability machine. In C. E. Shannon & J. McCarthy (Eds.), *Automata studies.* Princeton: Princeton Univer. Press, 1956. Pp. 277–285.

(Received March 27, 1961)

Psychological Review
1962, Vol. 69, No. 6, 501–519

PATTERNS OF ADRENERGIC-CHOLINERGIC IMBALANCE IN THE FUNCTIONAL PSYCHOSES

LEONARD S. RUBIN

Eastern Pennsylvania Psychiatric Institute, Philadelphia

During this last decade remarkable strides have been made in the treatment and control of the so-called functional psychoses with the aid of chemical agents. In view of the current status of the science of chemotherapy in this area, it is surprising that such beneficent results have been reported because there is no widely accepted chemical theory of the functional psychoses. A therapist is unable to prescribe, on the basis of preliminary laboratory examinations (of bodily fluids, for example), the chemical agent that is relevant to any specific endogenous chemical disequilibrium. One empirical chemotherapeutic approach consists of administering drugs on the basis of phenomenological impressions of simple behavioral dimensions such as depression or excitation. It need not be emphasized that few empirical findings and no theoretical formulations extant specifically relate these behavioral dimensions to any disturbed tissue system or state of chemical disequilibrium that is reversible by chemical intervention. In another approach, the tacit assumption is made by therapists administering drugs that the population of functionally psychotic individuals is homogeneous. In both approaches, when improvement does occur in some fraction of the population, it had not been predicted.

Prior to the formulation of several theories that have served to stimulate a great deal of recent research, numerous empirical biochemical and physiological studies were performed in the hope of uncovering some characteristic endogenous feature that differentiated normals from psychotics. The following observations are representative of a larger number of studies that suggest that the functional psychoses have as a concomitant some disorder of homeostasis. Fischer (1953) reported that schizophrenic blood was toxic to tadpoles, while Federoff and Hoffer (1956) reported it toxic to fibroblasts grown in tissue culture. The latter researchers, however, found the blood of patients undergoing surgery equally toxic. This suggested that a nonspecific perseverative response to stressful stimuli may be involved. In addition, various measurements of the response to stress have been made employing the following variables which reflect autonomic activity: pulse rate, blood pressure, skin circulation, and response to adrenalin (Finkelman & Stephens, 1936; Pfister, 1938). The results of these studies on chronic schizophrenics generally indicate that they are less reactive than healthy people. When a large series of measurements was made of basal metabolic rate, circulation rate, body temperature, excretion of 17-ketosteroids, thyroid activity, and water metabolism in schizophrenics and in healthy people, it was found that the variance of the measurements was greater for the schizophrenics (Richter, 1957). These studies again suggest that a concomitant of schizophrenia is some disorder of homeostasis.

Devoid of specific deductively elaborated hypotheses, this plethora of empirical findings contributed by physiologists, pharmacologists, and behavioral scientists stimulated the need for a rational approach to chemotherapy. Several astute attempts at systematization have been made. They will be described briefly because they serve to lay the groundwork for the major thesis contained in this report; that is, that discrete patterns of autonomic dysfunction may be etiological factors in the functional psychoses.

The major contemporary chemical theories of the functional psychoses are inextricably dependent upon the behavioral and biochemical effects of two psychotomimetic agents, mescaline and lysergic acid diethylamide (LSD-25) (Rubin, 1957). A biochemical theory of schizophrenia put forward by Osmond and Smythies (1952) and stated more specifically by Hoffer, Osmond, and Smythies (1954) postulated the existence of an endogenous mescaline-like substance, the "M substance," a toxic metabolite formed by the defective metabolism of epinephrine. They claimed that, whereas normal individuals metabolize epinephrine by amine oxidase or sulfoesterase, the psychotic metabolizes epinephrine by a phenolase. The resulting production of quinone indoles, i.e., adrenochrome and adrenolutine, interferes with cerebral metabolism. Hoffer (1957) administered adrenolutine (a more stable isomer of adrenochrome) and a placebo to a group of subjects and he attempted to determine which persons had been given adrenolutine and which the placebo. The results of the experiment were equivocal. Gastaldi, as reported in a review by Smythies (1958), evaluated the effect of the more highly reactive, unstable adrenochrome and reported the absence of behavioral abnormalities. In a recent study (Holland, Cohen, Goldenberg, Sha, & Leifer, 1958), equilibrium plasma concentrations of adrenaline and noradrenaline were measured in healthy people and schizophrenics. No significant difference was found in the rate of utilization of circulating adrenaline and noradrenaline. Axelrod (1958) attempted to detect the presence of adrenochrome in schizophrenics and found that it was absent in schizophrenics as well as in healthy people. As yet, this chemical theory of schizophrenia that postulates the existence of an endogenous metabolite of adrenaline produced by a deficiency in the process of methylation lacks empirical verification.

The above biochemical theory of schizophrenia was established following the observation that mescaline, an hallucinogen, was similar in chemical structure to the neurohumors, adrenaline and noradrenaline. The second theory to be considered was formulated after the observation was made that the most powerful hallucinogen, LSD, was a potent peripheral antagonist of serotonin (5-hydroxytryptamine). The clue to the important role of serotonin came from its discovery in the brain. Like acetylcholine, the substance is present in nervous tissue in a precursor state and is active only in an unbound form. The amine is unevenly distributed in the brain, its concentration being highest in the brainstem, especially the hypothalamus, lowest in the cortex, and almost undetectable in the cerebellum. The high biological activity of this amine strongly suggests an important function in chemical mediation in subcortical centers. Additional support for the role of serotonin in neural transmission comes from consideration of the distribution of the enzyme monoamine oxidase, the enzyme that destroys serotonin, and 5-hydroxytryptophan decarboxylase, the enzyme that synthesizes serotonin. Both enzymes are found in highest concentration in the hypothalamus. These considerations led Brodie and Shore (1957) to hypothesize that serotonin was an important chemical mediator of neural activity in the brain. Although the hypothesis has been fruitful, no current evidence has established beyond doubt that serotonin plays a transmitter role or is auxiliary to the action of other transmitters. Woolley and Shaw (1954) accepted the presumptive evidence for serotonin as a central neural transmitter substance and went on to postulate that the hallucinogenic effect of LSD might be due to its interference with the action of serotonin in the brain. According to these authors, the action of LSD, as well as other alkaloids, can be ascribed to the indole moiety in the structures of LSD and serotonin. Consequently, compounds containing the indole ring may act as antimetabolites to serotonin. This suggested to them that mental aberrations were produced by endogenously produced indole compounds that competed for serotonin receptor sites in the brain. This attractive theory has been challenged by at least two recent observations. Cerletti and Rothlin (1955) found that a close analogue of LSD, 2-bromo-D-lysegic acid diethylamide, was as effective as LSD in antagonizing several peripheral actions of serotonin, yet this antagonist of serotonin was found to be devoid of hallucinogenic action in humans. It should be noted that chlorpromazine is also a potent antiserotonin and, instead of being hallucinogenic, is a powerful antihallucinogen employed in the treatment of psychotics.

As indicated by the foregoing review, considerable effort is being expended to understand the biochemistry of both the LSD and the mescaline psychotic-like states in normal individuals. Although the evidence provided by researchers who have worked within these conceptual frameworks has been inconclusive, the quest for the specific toxic endogenous metabolite goes on. However, we still quite realistically speak of the chemical control of schizophrenia and not the chemical treatment.

A more detailed critical review of these theories led the author (Rubin, 1959) to conclude that it was far from ineluctable that they would ultimately lead to an effective chemotherapy for the functional psychoses. At this point, the author would introduce his theory by stating that the functional psychoses are characterized by patterns of *disordered* neurohumoral activity in the central representation of the autonomic nervous system. Since adjustments of the internal milieu are mediated by the autonomic nervous system and integrated in the hypothalamus, it is necessary to consider the humoral nature of hypothalamic activity before the theory is elaborated and supporting evidence is presented. Specifically, the assumption will be examined that the central representations of the autonomic nervous system are cholinergic for the parasympathetic division and adrenergic for the sympathetic division.

Vogt (1954) examined the concentration of sympathin (a combination of norepinephrine and epinephrine) in the different parts of the central nervous system of the dog and found that sympathin, like acetylcholine, possessed a specific pattern of distribution. Two regions of the brain, the hypothalamus and the area postrema, stood out above all others by their high concentrations of norepinephrine. This fact suggests, though it does not prove, that these amines play a part in the specialized function of those regions of the brain in which their concentration is high. Vogt's conclusion regarding the correlation between sympathin content and central representation of sympathetic functions was summarized in the following statement:

> With exception of the area postrema, those regions which contain the highest concentration of sympathin—the hypothalamus, parts of the midbrain and the floor of the fourth ventricle—also contain the central representation of sympathetic activity (p. 473).

However, when Vogt attempted to localize the site of sympathin within the hypothalamus, it was found that the anterior and posterior hypothalamus did not show consistent differences in sympathin content. This is of some interest because of the controversy over whether or not the anterior hypothalamus is the site of parasympathetic, and the posterior hypothalamus is the site of sympathetic centers. In this particular histochemical aspect, the two regions resemble each other. When the hypothalamus was divided into dorsal and ventral parts, however, a definite concentration gradient was found. The ventral region contained more noradrenaline than the dorsal region. Based upon Vogt's findings, the position taken in this paper is that the central representation of sympathetic function in the hypothalamus contains the highest concentration of the biogenic amines; norepinephrine, epinephrine, and serotonin.

In attempting to ascertain whether the central representation of the autonomic nervous system may be cholinergic for the parasympatheic division, the following facts must be considered. For purposes of description, the hypothalamus in man may conveniently be subdivided into four regions: the supraoptic middle region located above the optic chiasm and rostral to it, the tuberal or infundibular middle region, the mammilary middle region, and the lateral region. For a purpose that will become clearer as the discussion continues, the composition of the supraoptic middle region will be described. It includes the nuclei supraopticus, paraventricularis, suprachiasmaticus, supraopticus diffusus, and the anterior hypothalamic area (Kuntz, 1953). Feldberg and Vogt (1948) attempted to map out, in the central nervous system of the dog, the distribution of the enzyme (or enzyme system) that synthesizes acetylcholine. They found a very high value of enzymatic activity in the region of the supraoptic nuclei. By their definition, this portion of the hypothalamic area may be considered cholinergic. However, the high figures for the supraoptic nucleus do not suggest that cholinergic fibers originate from cells in this region, rather, such fibers terminate in synaptic connection with these cells. Unfortunately, the enzymatic activity of the anterior hypothalamus, the controversial site of parasympathetic activity, was not determined.

It seems reasonable to conclude from the foregoing discussion that specific areas within the hypothalamus cannot be viewed as exclusively adrenergic or cholinergic. The evidence does suggest that the hypothalamus contains large quantities of the biogenic amines as well as components of the cholinergic system. Consequently, in the remainder of this paper, when describing the central adrenergic-cholinergic imbalance that characterizes the functional psychoses, I do not necessarily implicate specific sympathetic and parasympathetic central structures as such.

Since Cannon's (1932) formulation of his Emergency Theory of Emotion, it has been accepted generally that, whenever an organism is confronted with a sudden emergency, the resulting massive autonomic response is functionally adaptive. Specifically, a massive and complete mobilization is achieved by the sympathetic division of the autonomic nervous system with the resultant secretion of great quantities of epinephrine. As Grenell states in Fulton (1949), "The so-called 'fight, flight or fright reaction' is a product of maximal generalized sympathetic activity." One major feature of an organism's response to an emergency situation, therefore, is a significant increase in the release of certain sympathomimetic amines. The ascription of a role to the parasympathetic division of the autonomic nervous system during the emergency reaction has been conspicuously neglected. Its role has been confined to the protection, conservation, and restoration of bodily resources. The hypothesis here offered is that the requisite *maximal* generalized adrenergic activity ensuing the perception of threat is facilitated by a decrease in cholinergic outflow. If we consider, as Cannon did, the functional utility of the increased adrenergic response to stress, then it follows that maximal mobilization can be most effectively attained under a condition in which central cholinergic outflow actually decreases below levels that characterize the organism during nonstressful steady states. It is conjectured that, were cholinergic outflow to increase or remain at the level of nonstressful states during the time the organism is mobilizing for an emergency, the net adrenergic response would be less than maximal. As the effectiveness of the organism's adjustment to emergency is directly related to maximal mobilization of bodily resources, it is hypothesized that maximal mobilization can only result from increased activity of central adrenergic centers and reduction of central cholinergic outflow. As a corollary, it is inferred that the patterns of autonomic dysfunction characterizing the functional psychoses consist of significant departures from the characteristic interactions of the adrenergic and cholinergic components of the normal individual at rest or under stress.

It is further postulated that neurohumoral imbalance in the functional psychoses may involve either or both neurohumoral systems and that pathological variation in either system is not unidirectional. Conceivably, either malfunction of the adrenergic mechanism or the cholinergic or both may lead to significant departures from that reaction requisite for maximal mobilization of the organism. A hypothetical matrix of varieties of neurohumoral imbalance has been constructed by reviewing the efficacy and mode of action of certain drugs already employed with varying success in the treatment and control of the functional psychoses. Patterns of cholinergic and adrenergic malfunction in the psychoses may thus be inferred from pharmacological and clinical observations. From the following evidence, it has been deduced that neurohumoral dysfunction may be characterized by either an excessive or deficient adrenergic response mechanism.

A DISORDERED ADRENERGIC RESPONSE MECHANISM IN THE FUNCTIONAL PSYCHOSES

Iproniazid (Marsilid), an amine oxidase antagonist that inhibits the oxidative deamination of epinephrine, has been used extensively in the treatment of mental depression. *In vivo* studies on animals with iproniazid indicate that it can produce a relatively rapid and large (threefold) rise in brain serotonin (Zeller, Barsky, & Berman, 1955). Before proceeding to the statement of an important inference derived from these observations, it would be well to evaluate the controversial role of monoamine oxidase in the metabolism of biogenic amines.

There are at least four potentially possible metabolic routes for the enzymatic conversions of norepinephrine, epinephrine, and closely related catecholamines: deamination by monoamine oxidase, oxidation to adrenochrome and related products by cytochrome oxidase and tyrosine oxidase, conjugation to the glucuronide or sulfate, or O-methylation by catechol-O-methyl transferase (Koelle, 1959). Though the importance of monoamine oxidase in the oxidation of catecholamines is controversial, it is well established that dopamine, a precursor of norepinephrine and epinephrine, is oxidized by monoamine oxidase at a reasonably rapid velocity. Furthermore, according to Bacq (1949), there is little possibility that the catecholamines are oxidized to indole derivatives by cytochrome oxidase *in vivo*. As far as the third metabolic pathway listed above is concerned, the results of a great deal of work suggest that free catcholamines and their immediate metabolites from other pathways are conjugated in considerable proportions. Finally, LaBrosse, Axelrod, and Kety (1958) have suggested that catechol-O-methyl transferase is the enzyme of major importance in the metabolism of circulating norepinephrine and epinephrine. They arrived at this conclusion after they observed that over 90% of a 1-mg. dose of dl—epinephrine given to normal human subjects by intravenous infusion—was recovered as the free or conjugated form of the primary product, metanephrine, or its subsequent oxidation product, 3-methoxy-4-hydroxymandelic acid. It is important to note that the authors still invoke a role for monoamine oxidase. It participates in the oxidation of metanephrine to the mandelic acid derivative. Pekkarinen, Castrin, and Israel (1958) present evidence suggesting that amine oxidase acts directly upon the catecholamines. This position is supported by the work of Shore (1958) who found increased amounts of norepinephrine in the brain and other organs of small animals after treatment with iproniazid. On the basis of his studies, he concluded that: "norepinephrine, as well as serotonin, is normally metabolized in vivo largely through the action of monoamine oxidase." He criticized much of the experimental evidence which suggested that monoamine oxidase is unimportant in catecholamine metabolism by emphasizing the erroneous use by others of *in vitro* tissue preparations. When monoamine oxidase was blocked *in vivo* by Marsilid, Shore could not find any alternate major pathway for the metabolism of norepinephrine. Evidence for or against each metabolic route cannot begin to be adequately described in this paper. A reasonable conclusion is that while monoamine oxidase may not be the sole enzyme system involved in the degradation of the catecholamines, it must be considered a significant part of the mechanism.

To return to the effects of iproniazid on the concentration of biogenic amines, it is likely, according to Truitt (1958), that this drug, through its inhibition of monoamine oxidase in the brain, potentiates the action therein of endogenous stimulant biogenic amines (possibly norepinephrine, epinephrine, and serotonin). The side effects (constipation, difficulty in micturition, hyperreflexia, sweating, and dryness of the mouth) are also compatible with a hypothesis of peripheral sympathetic stimulation. The implication drawn from these observations is that *some* psychotics who are benefited by iproniazid are defective in some brain catecholamines, or in central adrenergic stimulation.

When one turns to the treatment or control of the schizophrenic reaction, it is found that reserpine and chlorpromazine have been extensively used. Hyperemotional responses mediated by sympathetic centers in the hypothalamus are markedly depressed by the action of reserpine. Peripherally, the syndrome produced by reserpine is marked by bradycardia, slight lowering of blood pressure, stimulation of peristalsis, and gastric hypersecretion. Brodie and Shore (1957) proposed that reserpine prevented serotonin binding and thereby caused constant activation of the vegetative, parasympathetic centers. In addition, reserpine was found to deplete the store of biogenic catecholamines in the body as determined by the increased excretion of 5-hydroxy-indoleacetic acid, a metabolite of serotonin. Evidence has been presented in Truitt which favors a role for serotonin as a central excitatory agent at least in subcortical centers. The hyperactivity of serotonin, rather than its deficiency, therefore, may play a dominant role in mediating uncontrolled emotional responses in acute psychoses. Thus, the mechanism of reserpine action may be by depletion of brain concentration of serotonin and other catecholamines rather than through the sedative action of released serotonin. These observations therefore suggest that *some* functionally psychotic patients for whom reserpine therapy is efficacious may be characterized by an *excessive* adrenergic activity in subcortical centers. This same conclusion is also applicable when chlorpromazine is considered, for, in addition to its peripheral antihistaminic and anticholinergic effects, it has been shown to have an adrenergic blocking action.

In summary, it is postulated that the neurohumoral dysfunction that characterizes *some* psychotics may be due to either (*a*) a neurohumoral deficiency of some sympathomimetic amine or to some enzyme disorder such as monoamine oxidase overactivity in the adrenergic excitatory mechanism of the brain; or, (*b*) a neurohumoral overabundance of some sympathomimetic amine or to some enzyme disorder such as monoamine oxidase underactivity in the adrenergic mechanism of the brain.

A DISORDERED CHOLINERGIC RESPONSE MECHANISM IN THE FUNCTIONAL PSYCHOSES

There are suggestions that the cholinergic component of generalized autonomic activity is also an important parameter of the functional psychoses. Forrer (1956) has recommended the use of large doses of tertiary atropine in the treatment of manic-depressives in the manic phase. He claimed that the patients in the manic phase of manic-depressive psychosis responded remarkably well and immediately to this chemotherapy but he offered no mechanism which could possibly account for the effectiveness of this drug. As the pharmacological action of atropine is that of a cholinergic blocking agent, it may be inferred that *excessive* brain acetylcholine or deficient acetylcholinesterase with manifest excessive cholinergic outflow is important to the understanding of the chemical fundament in some psychotics. A strong theoretical statement which in part supports this inferred mechanism as to the efficacy of atropine therapy has been offered by Hoffer et al. (1954). He has claimed that excessive parasympathetic activity, with concomitant increases in the concentration of acetylcholine, is an essential feature of the schizophrenic process. In contrast to Forrer, however, Hoffer believes that cholinergic blocking agents as well as cholinergic compounds should aggravate the schizophrenic process and perhaps make psychotics worse.

Paradoxically, Pfeiffer, Jenney, Gallagher, Smith, Bevan, Killam, Killam, and Blackmore (1957) have reported

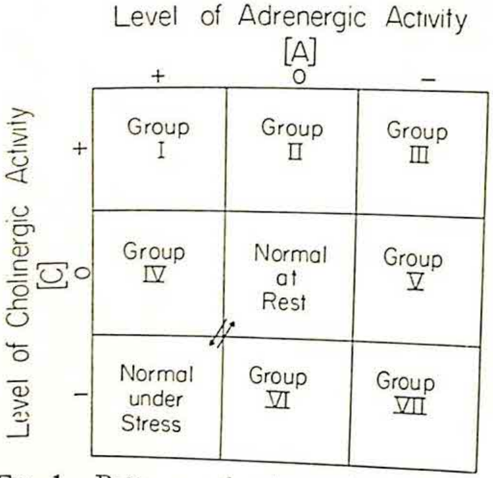


FIG. 1. Patterns of adrenergic-cholinergic interaction for normals and psychotics.

that some schizophrenics respond with a lucid interval to stimulation by cholinergic stimulants such as pilocarpine, arecoline, and 2-dimethylamine ethanol. The implication of this report is that some psychotics are characterized by *defective* cholinergic responsiveness. Recently, Rubin (1958a, 1958b) has reported that a functionally psychotic sample of patients was significantly differentiable from normals by virtue of the rate of hydrolysis of acetylcholine chloride by erythrocyte cholinesterase. All of the patients employed in the study manifested a defect in rate of hydrolysis. Some subjects hydrolyzed the substrate too rapidly and others too slowly. The inference made, on the basis of the kinetics of the enzymatic reaction, was that some psychotics may be deficient in cholinergic activity while others may be characterized by excessive cholinergic activity.

PATTERNS OF DISORDERED ADRENERGIC-CHOLINERGIC STEADY STATES IN PSYCHOTIC DISORDERS

A hurried review of the evidence presented seems replete with paradoxical facts and theories. At first it appears that an effective chemotherapy for some psychotics should counteract deficient adrenergic activity. Then, it seems that more effective drugs which decrease adrenergic activity should be developed. Additional conflict and confusion is raised by assertions of efficacy for cholinergic as well as anticholinergic drugs. These apparent contradictions may be resolved by regarding the population of functional psychotics as consisting of a specifiable number of discrete patterns of disordered adrenergic-cholinergic levels of reactivity. From the preceding review of the pharmacological action of certain drugs on the neurohumoral components of the autonomic nervous system, it is assumed that in the population of psychotics there may exist three levels of adrenergic activity: overreactivity, deficiency, and a normal level from which directional deviations are measurable. The cholinergic outflow may also be described as consisting of three levels: overreactivity or deficiency of cholinergic levels as measured by reference to a level characteristic of the unstressed normal. The matrix of adrenergic-cholinergic steady states derivable from such a paradigm is presented in Figure 1. In this figure, A represents central adrenergic activity and C central cholinergic activity. The levels of activity of each dimension are symbolized by +, 0 and −. The positive sign represents excessive activity, the zero, normal activity, and the negative sign, deficient activity. The combination of AoCo represents the levels of adrenergic and cholinergic activity that characterize the normal human adult at rest. The eight other combinations represent hypothetical adrenergic-cholinergic levels of interaction. Previously, the hypothesis had been developed that a normal, ordered response to stress consisted of maximal generalized sympathetic activity that was attained when cholinergic stimulation was decreased (C−), while adrenergic stimulation continued to increase (A+) up to that maximal level which characterized the individual. This characteristic adjustment is contained in Figure 1, and may be represented as A+C−. It is now further hypothesized, within the framework of the general theory, that the remaining seven interactions between adrenergic-cholinergic activity represent the discrete patterns which comprise the population of functional psychotics. To recapitulate, the major hypotheses elaborated from the above considerations are: *(a)* Effective autonomic adjustment to emergency situations requires a *maximal* generalized adrenergic response; *(b)* To attain a maximal level of sympathetic activity, it is not only necessary that increased levels of adrenergic activity follow the emergency but that the level of cholinergic activity decrease; *(c)* The functional psychoses are characterized by patterns of neurohumoral dysfunction that deviate significantly from the pattern of the normal, healthy individual at rest, as well as from the adrenergic increase and cholinergic decrease that characterizes the normal individual during stress; and *(d)* The varieties of neurohumoral dysfunction represent discrete levels of activity of interacting adrenergic and cholinergic mechanisms. That is, each pattern of neurohumoral dysfunction represents an impairment of function of either the adrenergic or cholinergic mechanism or both.

PUPILLARY REACTIVITY AS A METHOD FOR INVESTIGATING PATTERNS OF DISORDERED ADRENERGIC-CHOLINERGIC STEADY STATES IN PSYCHOTIC DISORDERS

The hypotheses elucidated would remain interesting speculations unless a method were described which could systematically measure the relevant parameters. To this end a physiological measure of adrenergic and cholinergic activity was developed for investigating the patterns of disordered autonomic steady states in psychotic disorders. The miotic response of the pupil to light (constriction) and the mydriasis ensuing the absence of light (dilatation) were employed to measure cholinergic and adrenergic mechanisms, respectively. The rationale of the procedure is suggested by the following considerations. The cholinergic nature of the activity of the oculomotor nerve under the influence of light has been proved by the positive assay for acetylcholine in the anterior chamber of the eye (Engelhart, 1931) and by the disappearance of the light reflex after the injection of cholinesterase (Mendel & Hawkins, 1943). Conversely, the anticholinesterases induce miosis, presumably by interfering with the hydrolysis of acetylcholine by acetylcholinesterase. The adrenergic nature of the dilatory response of the pupil is suggested by the work of Poos (1939), Hess and Koella (1950), and Hess, Koella, & Szabo (1950) on isolated radial preparations of the iris. These studies demonstrated that the iris muscle contracted in response to small concentrations of adrenalin and noradrenalin. Since these substances are responsible for neurohumoral transmission in sympathetically innervated structures, they believed it reasonable to assume that a dilator muscle existed that responded to sympathetic excitation. In addition, Frohlich and Loewi (1910) had shown that the mydriasis induced by adrenalin is intensified by cocaine, a monoamine oxidase inhibitor.

Since pupillary constriction had been shown to be dependent upon the liberation of acetylcholine, and dilatation upon an adrenergic mediator, the rate of dilatation was considered to represent the kinetics of a general adrenergic mechanism, while the rate of constriction was considered indicative of the kinetics of the cholinergic mechanism. If the varieties of disordered neurohumoral function that characterize the functional psychoses do represent discrete levels of activity of interacting cholinergic and adrenergic mechanisms that differ significantly from the patterns manifested by normals at rest and under stress, then different patterns of constriction and dilatation should be observable. That is, it was assumed that peripheral cholinergic and adrenergic imbalances, respectively, mirror central cholinergic and adrenergic imbalances.

PROCEDURE

Although the apparatus and procedure for obtaining measurements of pupillary dilatation and constriction have been described elsewhere (Rubin, 1960, 1961a, 1961b), it seems appropriate to emphasize certain features of the procedure for increased clarity in exposition. To obtain the rate of dilatation for each subject, the following procedure was employed. After the subject was seated with his head positioned properly in a head rest, the room light was extinguished. After 5 seconds in the dark, an intense adapting light of 38.6 ft.-C. was focused on the subject's right eye for 10 seconds. At the termination of this interval, a photograph of the maximally constricted pupil was taken automatically. This was done twice in an effort to equate for previous conditions of adaptation of the retina. Each subsequent photograph of the pupil was taken after the subject had been dark-adapted for a prescribed interval of time following maximal constriction. The intervals of dark-adaptation employed were 5, 10, 15, 30, 45, and 60 seconds. The procedure employed to obtain measurements of the rate of pupillary constriction was initiated soon after measurements of pupillary dilatation had been obtained. Each photograph of the pupil was taken after the subject had been light-adapted for a specific interval of time following maximal dilatation. The intervals of light-adaptation employed were 1, 2, 3, 4, and 5 seconds. Prior to obtaining a measurement of constriction at each of these intervals, the eye was light-adapted for 10 seconds at a constant intensity before it was permitted to dilate maximally within the 60-second dark-adaptation period.

PATTERN OF ADRENERGIC-CHOLINERGIC ACTIVITY OF THE NORMAL ADULT UNDER STRESS

To test the hypothesis that normal adult subjects react to emergencies with an increased adrenergic response and decreased cholinergic response, measurements of pupillary dilatation and constriction were obtained on six normal, healthy adults between the ages of 19–40 (Rubin, 1960). Due to the variety of perceptual responses that are elicited in humans to the same stimulus, it is difficult to define a stimulus situation that is realizable in a laboratory as an "emergency" or as a "threat." Consequently, following the control measurements obtained when the subjects were at rest, each subject was required to immerse the right hand and forearm in a cold water bath maintained at 5° C. while measurements were made of dilatation and constriction. When the rate of pupillary dilatation at rest was compared to the rate obtained during the stress condition, it was found that the slopes of the curves were equal (parallel) but that the absolute level of dilatation under the cold pressor condition was significantly greater (Figure 2). Assuming that the responsiveness of the dilator pupillae is an increasing monotonic function of the amount of adrenergic mediator liberated at the effector site, it may be concluded that under stress a greater quantity of the adrenergic mediator is liberated. Therefore, the finding that the cold pressor test increased pupillary dilatation significantly was interpreted as supporting

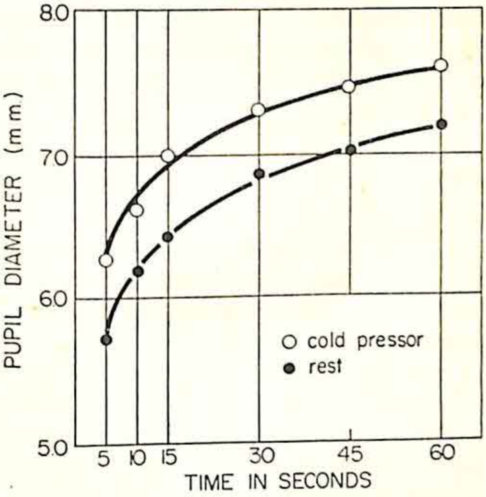


FIG. 2. Pupillary dilatation in normal subjects exposed to the cold pressor test.

the hypothesis that under a condition of stress, the normal human adult responds with an increased adrenergic outflow.

When the rates of pupillary constriction over a period of 5 seconds for the normal individuals at rest and under the condition of the cold pressor were compared, a significant difference was obtained (Figure 3). As opposed to the relationship observed during dilatation, the level of constriction of the pupil under the condition of the cold pressor test was decreased. This suggested a decreased cholinergic outflow in normal humans under stress, when it was assumed that the degree of constriction of the sphincter pupillae was an increasing monotonic function of the amount of acetylcholine liberated at the effector sites.

The significance of these findings, however, is contingent upon two basic assumptions: that the cold pressor test is representative or equivalent to other forms of physiological or psychological stress, and that the adrenergic and cholinergic reactivity of the pupil is representative of adrenergic-cholinergic mechanisms of interaction in vital neural integrative centers. If these assumptions are accepted, the following interpretation of the results of the experiment may be made with reasonable assurance. Upon exposure to a stress-engendering situation, the normal human organism responds with an increased adrenergic outflow and a decrement in cholinergic outflow. This reciprocal balance between the systems insures maximal mobilization of the resources of the organism required for "fight, flight, or fright."

ADRENERGIC-CHOLINERGIC REACTIVITY LEVELS OF FUNCTIONAL PSYCHOTICS

Several additional experiments were performed by the author (Rubin, 1961a, 1961b) in order to test two additional hypotheses. The first hypothesis tested was that the population of psychotics exhibited three distinct levels of adrenergic activity and three distinct levels of cholinergic activity. The second hypothesis tested was that psychotics manifest patterns of neuro-

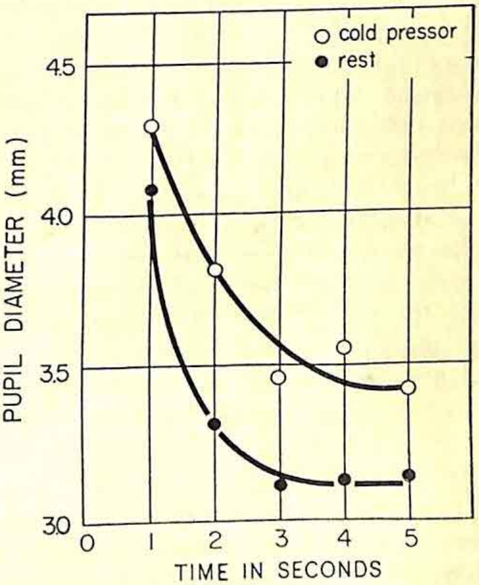


FIG. 3. Pupillary constriction in normal subjects exposed to the cold pressor test.

humoral dysfunction that deviate significantly from the pattern of normals at rest, as well as from the adrenergic increase and cholinergic decrease that characterizes the normal person during stress. Furthermore, the varieties of neurohumoral dysfunction could be represented by discrete patterns of levels of adrenergic-cholinergic interaction, as described previously in Figure 1.

To test these hypotheses, measurements of dilatation and constriction were obtained on 23 healthy, control subjects between the ages of 19–40 years. To determine the extent of dilatation between the fifth and sixtieth seconds, independent of the individual differences in the final level of dilatation attained, the following measurement was employed. The initial level of dilatation (mm.) was subtracted from the final level (mm.) and the difference was divided by the final level of dilatation. The ratio obtained in this manner for each subject represented the increase in diameter of the dark adapted pupil as a function of the maximal dilatation attained at the end of 60 seconds spent in darkness. Assuming that the sample of healthy subjects employed in the experiments is a random sample from a population with a normal distribution for the obtained ratio, then the sample mean, as an unbiased estimate of the population mean, may be employed to determine the likelihood of obtaining values for the ratio that deviate significantly in the direction of excessive or deficient dilatation. A graphic test to determine the normality of the distribution (Bliss & Calhoun, 1954) of the dilatation ratio was performed. The rankits were plotted against their corresponding variates and a straight line was fitted easily by inspection. This strongly suggested that the sample of dilatation ratio scores was distributed normally. One could now make inductive inferences from the sample with a reasonable degree of assurance. Consequently, the cumulative distribution function was employed to determine the relative frequency with which observations fell beyond the extreme dilatation ratios obtained from the sample.

In order to determine the extent of pupillary constriction, measurements of the diameter of the 1 minute dark-adapted pupil were obtained following exposure to a constant intensity of light that lasted for 1, 2, 3, 4, or 5 seconds. A constriction ratio for each subject was obtained in the following manner. To avoid negative values, the final level of constriction (mm.) was subtracted from the invariably greater initial level (mm.) and then divided by the final level. The graphic test for normality of the distribution of the ratios was performed and it was found that they were distributed normally. The cumulative distribution function was again employed to determine the likelihood of obtaining constriction ratios that fell beyond the extreme constriction values of the sample.

The probability of obtaining dilatation or constriction ratios significantly greater than or less than the extreme values in each distribution was then employed to evaluate the measurements obtained from a sample of psychotics.

The results to be described were obtained on two samples of patients who were tested at two different psychiatric institutions. As the results of the first experiment were confirmed by the second experiment, the following description pertains to 71 patients who had been diagnosed as functionally psychotic (Rubin, 1961b). Employing the probability limits obtained for the distribution of the dilatation ratio in the sample of normal individuals, it was found that 21 of 71 patients manifested excessive dilatation, 26 of 71 patients manifested deficient dilatation, and 24 of 71 patients manifested normal dilatation. These relationships are presented in Figure 4 where the mean dilatation ratio for each of the three groups at each time interval is presented as a function of the amount of time spent in darkness.

Inspection of the figure clearly indicates that there are three levels of dilatation that characterize the sample of patients employed in the experiment. The hypothesized three distinct levels of adrenergic activity, as measured by pupillary reactivity, seem to have been confirmed. That is, if it is assumed that the magnitude of pupillary dilatation is proportional to the quantity of adrenergic mediator liberated, then it may be concluded that some of the patients liberated a quantity that was characteristic of normals at rest; others liberated more of the adrenergic mediator or had some enzyme disorder such as monoamine oxidase underactivity; and some liberated too little of the adrenergic mediator or had some enzyme disorder such as overactivity of monoamine oxidase.

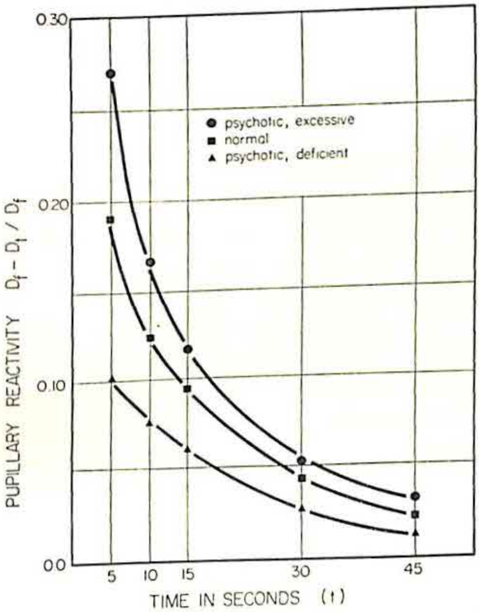


FIG. 4. Reactivity levels of pupillary dilatation in psychotic patients. (The values of the ordinate, D_f–D_t/D_f, consist of the final level of dilatation, mm., after 60 seconds of dark adaption, D_f, minus the level of dilatation at each interval in darkness, D_t, divided by the final level of dilatation, D_f. The abscissa represents the intervals of darkness in seconds, t, preceding photographing of the pupil.)

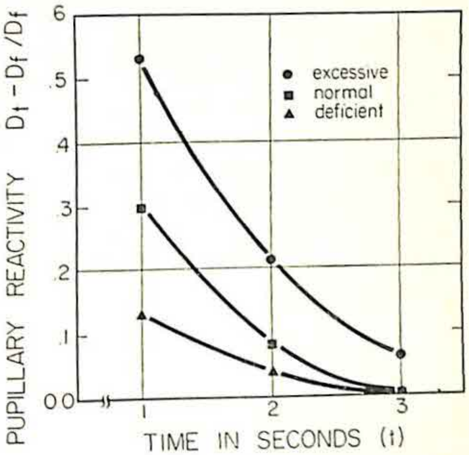


FIG. 5. Reactivity levels of pupillary constriction in psychotic patients. (The values on the ordinate, D_t–D_f/D_f, consist of the average diameter—mm.—of the constricted pupil at each interval, D_t, minus the diameter of the maximally constricted pupil after 5 seconds of light adaptation, D_f, divided by D_f. The abscissa represents the duration of light adaptation in seconds, t.)

Employing the probability limits obtained from the distribution of the constriction ratio in the sample of normal individuals, it was found that 36 of 71 patients manifested excessive constriction, 7 of 71 patients manifested deficient constriction, and 28 of 71 patients showed normal constriction. These relationships are presented in Figure 5 where the average constriction ratio for each of the three groups at each interval is represented as a function of length of time of exposure to a constant intensity light source. When the curve for those showing defective constriction is compared to the curve representing normal constriction, it is seen that the amount of change in pupil size for the deficient group remains smaller than for normal constriction. For the group manifesting excessive constriction, it is apparent that the amount of constriction per unit time is much greater than that of the normal or defective group. With respect to constriction, therefore, three levels, each significantly different from the other, were found in the sample of patients employed in the studies. Assuming that the tendency toward constriction of the sphincter pupillae is proportional to the quantity of acetylcholine liberated, the demonstration of three levels of constriction suggests that: (*a*) in one group of patients, either excessive quantities of acetylcholine were liberated or there was deficient acetylcholinesterase activity; (*b*) in a second group of patients, insufficient quantities of the neurohumor were liberated or enzyme activity was excessive; (*c*) the third group manifested that level of cholinergic activity that was equivalent to the level of normal subjects. These observations on the kinetics of adrenergic and cholinergic activity confirmed two hypotheses deduced from the theory; that is, the existence of three distinct levels of adrenergic activity and three distinct levels of cholinergic activity.

In order to establish that there are varieties of neurohumoral dysfunction that comprise the population of functional psychotics, it remained necessary to show that the disorders consisted of an impairment of function of either the adrenergic or cholinergic mechanism, or both. The demonstration of these patterns of interaction was derived from the following considerations. Twenty-four out of 71 psychotic patients, it will be recalled, showed no impairment in the extent of dilatation. However, *all* of the members in this group manifested *aberrant* constriction. Twenty of the 24 manifested excessive constriction, while 4 patients were deficient in constriction. Conversely, of the 28 psychotic patients out of 71 who showed no impairment in constriction, *all* were aberrant in the extent of dilatation; 12 showed excessive dilatation and 16 manifested deficient dilatation. In confirmation of the major hypothesis, it may be stated that all of the 71 patients employed in the studies showed a significant impairment in either the tendency to dilatation or to constriction or both. Furthermore, if the magnitude of dilatation is considered to be proportional to the quantity of adrenergic mediator liberated, and the amount of constriction of the sphincter pupillae proportional to the quantity of acetylcholine liberated, then it may be concluded that all of the psychotics employed in these studies were differentiable into seven discrete patterns of adrenergic-cholinergic imbalance.

The varieties of adrengeric-cholinergic dysfunction elucidated indicate that the population of functionally psychotic individuals is heterogeneous. Thus, the current quests for a chemical panacea that affects only one of the components of the autonomic nervous system undirectionally do not seem promising. Instead, the results of these studies suggest that each pattern of adrenergic-cholinergic activity may require a specific combination of drugs which differentially affect the level of activity of either or both components (Table 1). The value of the present study is that it seems to provide a rationale for the selection of patients who require distinct chemotherapies as determined by their specific pattern of adrenergic-cholinergic imbalance.

An obvious but unwarranted implication that may be drawn by the reader of the foregoing assertion is that the author believes with warrantable assertability that these patterns of autonomic dysfunction are etiological factors in the functional psychoses. Admittedly, it is difficult to conceive of an environmental process that could produce central neurohumoral dysfunction. However, the experiments described in this report were not designed to answer the question, "Is central neurohumoral dysfunction the invariable antecedent of psychotic behavior or the effect of dynamic, environmental variables?" Although the answer to the question may be of academic interest, it would be more fruitful to accept the attitude of a pragmatist and insist that truth is preeminently to be tested by the practical consequences of belief.

ADRENERGIC-CHOLINERGIC DYSFUNCTION IN PSYCHOTIC CHILDREN

The technique of employing pupillary measurements to determine adrenergic-cholinergic steady-states that characterize adult psychotics was also applied to ascertain whether autistic, psychotic children were significantly differentiable from normal children. An additional comparison could readily be made between those steady-states that characterized normal children and the adrenergic-cholinergic responsivity of normal adults. Thus an opportunity was also afforded to contribute to an understanding of the ontology of autonomic reactivity in normal individuals.

In the preliminary experiment to be described, data on dilatation and constriction was obtained on four normal, healthy children between the ages of 7–9.5 years, and five autistic children between the ages of 7–12 years. A statistical analysis of the rate of dilatation between these groups indicated that the slopes were equal and parallel but separated significantly by a constant (Figure 6). Again, assuming that the magnitude of the response of the dilator pupillae is an increasing monotonic function of the amount of adrenergic mediator liberated at the effector site, then the absolute level of adrenergic activity for the normal children is greater than that for the psychotic children. This suggests that normal children have a significantly greater capacity for increased adrenergic outflow than autistic children. When the two groups of children were compared with respect to the rate of constriction, it was found that the rate was significantly faster for the normal children, although no significant differences obtained in the final, absolute level of constriction (Figure 7). This result was interpreted as indicating that the rate of cholinergic activity was more rapid for normal

TABLE 1

A CLASSIFICATION OF FUNCTIONAL PSYCHOSES INTO DISCRETE LEVELS OF ADRENERGIC-CHOLINERGIC INTERACTION AND THE DERIVED CHEMOTHERAPIES

Group	Adrenergic-Cholinergic reactivity pattern	Suggested chemotherapy
I	A+C+	Adrenolytic agent in addition to an anticholinergic compound
II	AoC+	Anticholinergic compound
III	A−C+	Adrenergic stimulant in addition to an anticholinergic compound
IV	A+Co	Adrenolytic agent
V	A−Co	Adrenergic stimulant
VI	AoC−	Cholinergic stimulant
VII	A−C−	Adrenergic stimulant in addition to cholinergic stimulant

children. These results suggest that normal children have a greater capacity for sympathicoadrenal response, and, of equal importance, a greater manifest capacity to reestablish more quickly the vegetative functions mediated by the cholinergic system. Clearly, autistic children demonstrate a pattern of autonomic dysfunction when compared to normal children. However, the limited sample sizes employed in this preliminary experiment precluded the ultilization of the statistical techniques employed to establish discrete patterns of neurohumoral dysfunction in adult psychotics.

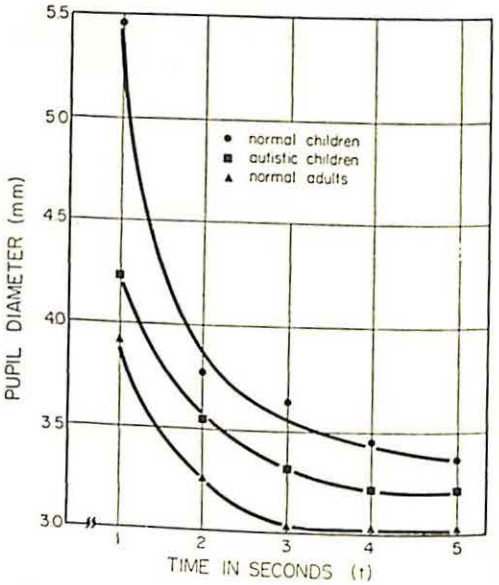


FIG. 7. Change in pupillary constriction as a function of duration of light adaptation for normal adults, normal children, and autistic, psychotic children.

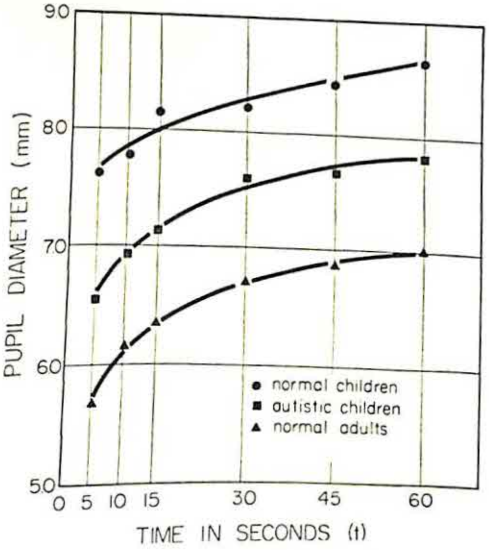


FIG. 6. Change in pupillary dilatation as a function of the amount of time spent in darkness for normal adults, normal children, and autistic, psychotic children.

Reference to Figure 6 reveals another experimental comparison that may throw some light upon the ontology of adrenergic responsivity. When the course of dilatation for normal children was compared to that of normal adults, it was found that the mean pupil diameter at each interval for the children was significantly greater than that of the normal adults. Again, if it is assumed that the amount of dilatation is proportional to the amount of adrenergic mediator liberated, then it may be concluded that normal children have a significantly greater capacity for increased adrenergic outflow. An interesting difference in cholinergic reactivity was revealed when the course of constriction in normal adults was compared to that of normal children. A statistical analysis of the results presented in Figure 7 revealed that normal adults manifested significantly greater constriction than normal children. However, although the absolute level of cholinergic activity was found to be consistently greater for the normal adult, the rate of cholinergic activity was found to be significantly greater for the normal child. These results suggest that normal children and normal adults are significantly different with respect to autonomic responsivity. Furthermore, although the intermediate biochemical (genetic?) factors responsible for the change in neurohumoral activity are not as yet known, it would seem that the normal child, as he develops into a normal adult, undergoes changes that reduce the level of adrenergic outflow, decrease the rate of cholinergic activity, and increase the absolute level of cholinergic outflow.

SUMMARY

The results of the studies described in this report confirmed several hypotheses derived from a theoretical formulation of the role of adrenergic and cholinergic mechanisms in psychotic disorders. The salient features of the theory are: effective adjustments to emergency situations require a maximal, generalized sympathetic response; to attain a maximal level of sympathetic activity, it is not only necessary that increased levels of adrenergic outflow follow the emergency but that the level of cholinergic activity decrease; the neurohumoral dysfunctions that characterize the psychoses represent significant departures from the adrenergic-cholinergic states that characterize the normal individual at rest and during stress; and the varieties of dysfunction represent discrete levels of activity of interacting adrenergic and cholinergic mechanism. That is, each pattern of neurohumoral imbalance represents an impairment of function of either the adrenergic or cholinergic mechanism or both.

The nature of the adrenergic impairment that characterizes some psychotics may be a neurohumoral deficiency of some biogenic amine or some enzyme disorder such as monoamine oxidase overactivity in the adrenergic centers of the brain. Other psychotics manifest adrenergic impairment by a neurohumoral excess of some sympathomimetic amine that may be the result of some enzyme disorder such as monoamine oxidase underactivity in central adrenergic centers. Similarly, the cholinergic impairment of some psychotics may be characterized by excessive concentrations of brain acetylcholine or deficient acetylcholinesterase, while other psychotics may possess an inadequate level of acetylcholine or excessive acetylcholinesterase activity. A theoretical matrix of neurohumoral steady states, derived from considerations of the interaction of the levels of the two systems, characterized the population of functional psychotics as well as the normal individual at rest and under stress.

The application of measurements of pupillary reactivity to the study of autonomic dysfunction in autistic, psychotic children demonstrated that adrenergic-cholinergic imbalance clearly differentiated these children from normal children. The significance of this obtained difference was evaluated with respect to the pattern of autonomic function manifested by normal adults.

That the adult psychotic may be classified into any one of seven discrete categories of adrenergic-cholinergic activity levels throws reasonable doubt upon the current quest for a single chemical panacea. Instead, the elucidation of these discrete patterns of adrenergic-cholinergic activity suggests a rational approach to chemotherapy. Each pattern of adrenergic-cholinergic dysfunction requires a specific combination of drugs that differentially affect the levels of activity of either or both components of the autonomic nervous system.

REFERENCES

AXELROD, J. Is adrenochrome present in the blood? *Amer. J. Psychiat.*, 1958, **115**, 162–163.

BACQ, Z. M. The metabolism of adrenaline. *Pharmacol. Rev.*, 1949, **1**, 1–26.

BLISS, C. I., & CALHOUN, D. W. *Outline of biometry*. New Haven, Conn.: Yale Cooperative Corporation, 1954.

BRODIE, B. B., & SHORE, P. A. A concept for a role of serotonin and norepinephrine as chemical mediators in the brain. *Ann. N. Y. Acad. Sci.*, 1957, **66**, 631–642.

CANNON, W. B. *The wisdom of the body.* New York: Norton, 1932.

CERLETTI, A., & ROTHLIN, E. Role of 5-hydroxytryptamine in mental diseases and its antagonism to lysergic acid derivates. *Nature, London,* 1955, **176**, 785–786.

ENGELHART, E. Der humorale Wirkungsmechanismus der Okulomotoriusreizung. *Arch. ges. Physiol.*, 1931, **227**, 220–234.

FEDEROFF, S., & HOFFER, A. Toxicity of blood serum from schizophrenics and nonschizophrenic subjects. *J. nerv. ment. Dis.*, 1956, **124**, 396–398.

FELDBERG, W., & VOGT, M. Acetylcholine synthesis in different regions of the central nervous system. *J. Physiol.*, 1948, **107**, 372–381.

FINKELMAN, I., & STEPHENS, W. M. Head regulation in dementia praecox; reactions of patients with dementia praecox to cold. *J. Neurol. Psychopath.*, 1936, **16**, 321–340.

FISCHER, R. Stress and toxicity of schizophrenic serum. *Science,* 1953, **118**, 409–411.

FORRER, G. R. Symposium on atropine toxicity therapy: History and future research. *J. nerv. ment. Dis.*, 1956, **124**, 256–259.

FROHLICH, A., & LOEWI, O. Über eine Steigerung der adrenalin Empfindlichkeit durch Cocain. *Arch. exp. Path. Pharmakol.*, 1910, **62**, 159–169.

FULTON, J. F. *A textbook of physiology.* (16th ed.) Philadelphia: Saunders, 1949.

HESS, W. R., & KOELLA, W. Experimentelle Studien über die antagonistische Innervation. *Z. ges. exp. Med.*, 1950, **116**, 431–443.

HESS, W. R., KOELLA, W., & SZABO, T. Experimentelle Studien über die antagonistische Innervation. *Z. ges. exp. Med.*, 1950, **115**, 481–490.

HOFFER, A. Epinephrine derivatives as potential schizophrenic factors. *J. clin. exp. Psychopath.*, 1957, **18**, 27–60.

HOFFER, A., OSMOND, H., & SMYTHIES, J: Schizophrenia: New approach; result of year's research. *J. ment. Sci.*, 1954, **100**, 29–54.

HOLLAND, R., COHEN, G., GOLDENBERG, M., SHA, J., & LEIFER, I. Adrenaline and noradrenaline in urine and plasma of schizophrenics. *Federat. Proc.*, 1958, **17**, 378.

KOELLE, G. B. Possible mechanisms for the termination of the physiological actions of catecholamines. *Pharmacol. Rev.*, 1959, **11**, 381, 386.

KUNTZ, A. *The autonomic nervous system.* (4th ed.) Philadelphia: Lea & Febiger, 1953.

LABROSSE, E., AXELROD, J., & KETY, S. O-methylation, the principal route of metabolism of epinephrine in man. *Science,* 1958, **128**, 593–594.

MENDEL, B., & HAWKINS, R. D. Removal of acetylcholine by cholinesterase injections and effect thereof on nerve impulse transmission. *J. Neurophysiol.*, 1943, **6**, 431–438.

OSMOND, H., & SMYTHIES, J. Schizophrenia: New approach. *J. ment. Sci.*, 1952, **98**, 309–315.

PEKKARINEN, A., GASTREN, O., & IISALO, E. Effect of iproniazid on ganglionic-blocking substances on the noradrenaline content of tissues. In, *Fourth International Congress of Biochemistry,* 1958, **15**, 106–107.

PFEIFFER, C. C., JENNEY, E. H., GALLAGHER, W., SMITH, R. P., BEVAN, W., KILLAM, K. F., KILLAM, E. E., & BLACKMORE, W. Stimulant effect of 2-dimethyl aminoethanol. *Science,* 1957, **126**, 610-611.

PFISTER, H. D. Disturbances of autonomic nervous system in schizophrenia and their relations to insulin cardiazol and sleep treatments. *Amer. J. Psychiat.*, 1938, **94**, 109–118.

POOS, F. Über die Eignung der Pupillae als Testobjekt für pharmacologische Reaktionen und Pharmakodiagnostik an Auge. *Ergbn. Physiol.*, 1939, **41**, 883–916.

RICHTER, D. (Ed.) *Schizophrenia: Somatic aspects.* New York: Macmillan, 1957.

RUBIN, L. S. The psychopharmacology of lysergic acid diethylamide (LSD-25). *Psychol. Bull.*, 1957, **54**, 479–489.

RUBIN, L. S. Acetylcholine hydrolysis in psychiatric patients. *Science,* 1958, **128**, 254–255. (a)

RUBIN, L. S. Cholinesteraese activity. *Science,* 1958, **128**, 1176–1178. (b)

RUBIN, L. S. Recent advances in the chemistry of psychotic disorders. *Psychol. Bull.*, 1959, **56**, 375–383.

RUBIN, L. S. Pupillary reactivity as a measure of adrenergic-cholinergic mechanisms in the study of psychotic behavior. *J. nerv. ment. Dis.*, 1960, **130**, 386–400.

RUBIN, L. S. Autonomic dysfunction in psychotic adults and autistic children. *Arch. gen. Psychiat.*, 1962, **7**, 1–14. (a)

RUBIN, L. S. Patterns of pupillary dilatation and constriction in psychotic adults and autistic children. *J. nerv. ment. Dis.*, 1961, **133**, 130–142. (b)

SHORE, P. A. Possible mechanism of antidepressant action of Marsilid. *J. clin. exp. Psychopath.*, 1958, **19**, 56–60.

SMYTHIES, J. R. Biochemical concepts of schizophrenia. *Lancet,* 1958, **2**, 308–313.

TRUITT, E. B. Some pharmacologic correlations of the chemotherapy of mental disease. *J. nerv. ment. Dis.*, 1958, **216**, 184–210.

VOGT, M. The concentation of sympathin in different parts of the central nervous system under normal conditions and after the administration of drugs. *J. Physiol.*, 1954, **123**, 451–481.

WOOLLEY, D. W., & SHAW, E. A biochemical and pharmacological suggestion about certain mental disorders. *Science,* 1954, **119**, 587–588.

ZELLER, E. A., BARSKY, J., & BERMAN, E. R. Amine oxidases inhibition of monoamine oxidase by 1-isonicotinyl-2-isoprophylhydrazine. *J. biol. Chem.*, 1955, **214**, 267–274.

(Received June 27, 1961)

Psychological Review
1962, Vol. 69, No. 6, 520–536

SUCCESS AND SPEED OF PROBLEM SOLVING BY INDIVIDUALS AND GROUPS[1]

FRANK RESTLE[2] AND JAMES H. DAVIS[3]

Michigan State University

The problem solving process, which begins with the presentation of a problem, is terminated when the subject arrives at a correct solution, when he arrives at an incorrect solution he believes correct, or when the experimenter arbitrarily ends the experimental session. Performance on a problem is usually summarized by the proportion of subjects who solve, or by the mean or median time to solution. A more complete description of the data gives the success of each subject (whether he obtained the correct answer or not) and the time taken; that is, a distribution of solution times or times until terminal error. One purpose of this paper is to show how the full distributions can be used to answer theoretical questions about the structure of the problem.

Small ad hoc groups of subjects can work together on verbal puzzles. Such cooperation seems natural and is appropriate to the nature of the problem and the usual behavior of college students. Generally speaking, small groups solve problems more rapidly and surely than individuals from the same population (Faust, 1959; Husband, 1940; Lorge, Fox, Davitz, & Brenner, 1958; Lorge, & Solomon, 1955; Marquart, 1955; Shaw, 1932; Taylor & Faust, 1952; Thorndike, 1938; and Watson, 1928). The advantage of groups may be a simple effect of pooling, "two heads are better than one," for in cooperative problem solving the group gets credit for the first individual insight. Other factors such as interference between group members or emergent group facilitation of thinking must be measured as deviations from the simple "pooling" hypothesis. The second aim of this paper is to formulate the pooling hypothesis so that it applies to distributions of solution times.

MODEL OF THE DISTRIBUTION OF SOLUTION TIMES

Suppose that time is divided into short equal intervals of duration h. In any interval the subject may either solve the problem or not. Imagine that there exists a probability φ that the problem is solved in the nth interval given that it has not been solved previously. The probability φ is taken by hypothesis to be constant over intervals. Then the probability that the problem is actually solved in exactly the nth interval is the probability that it is not solved in any of the first $n-1$ intervals (an event with probability $(1-\varphi)^{n-1}$) times the probability that it is solved in the nth interval. That is,

$$P(n) = \varphi(1-\varphi)^{n-1} \qquad [1]$$

the geometric distribution. This is a positively skewed (J-shaped) distribution with mean $1/\varphi$ and variance $(1 - \varphi)/\varphi^2$.

---

[1] Developed from part of a doctoral dissertation submitted to Michigan State University by James H. Davis (1961). The first author directed the dissertation problem.

[2] Faculty Research Fellow of the Social Science Research Council. Now at Indiana University.

[3] During part of the work of this problem, Summer Fellow of the National Science Foundation. Now at Miami University (Ohio).

To say that φ is constant over trials, that the probability of solving is independent of how many times the subject may have failed to solve the problem, is in effect to say that problem solving is a one-stage, all-or-nothing process. This assertion is frequently unreasonable in that some problems clearly seem to involve more than one stage.

Multi-stage Problems

Suppose that to obtain the final solution the subject must actually solve a sequence of stages. The stages may be solved in fixed order, or it may be that the order of stages can be inverted, but suppose that the subject actually can work on only one stage at a time. Furthermore, suppose that the time to solve an individual stage has the geometric distribution, and that the distribution of times to solve Stage 2 is independent of how long it took to solve Stage 1, etc. (Of course, it is assumed that work on Stage 2 does not even start until Stage 1 is completed.) The distribution of times to solution for a k stage problem can be determined; it is a random variable, and the sum of the k random variables which are the times to complete the k stages. If the k stages have the same probability of solution per interval, φ, then the probability that the kth stage (and the whole problem) are completed at just trial N, is given by the negative binomial (Pascal) distribution,

$$P(N; k, \varphi) = \binom{N-1}{k-1} \varphi^k (1 - \varphi)^{N-k} \quad [2]$$

An intuitive justification of this formula is given by Bush and Mosteller (1955), Ch. 14) and its derivation is discussed by Feller (1957, pp. 155–157). Note that Equation 2 reduces to Equation 1 when $k = 1$. The expected number of time intervals to solution, the expectation $\mathrm{E}(N)$, and the variance of number of intervals, are given by

$$\mathrm{E}(N) = k/\varphi \quad [3]$$

$$\mathrm{Var}(N) = k(1 - \varphi)/\varphi^2 \quad [4]$$

Although it is not likely that the various stages of a given problem would be equally difficult (have equal values of φ), there is great difficulty in estimating φ's of separate stages unless one has direct observations of the completion of various stages. Such analyses can sometimes be done but it is helpful to be able to deal directly with total solution time. For this purpose it is expedient to use the simplifying hypothesis that all stages are equally difficult, so as to be able to estimate the number of stages in the problem.

Continuous Time

The above remarks deal with short intervals of time, yet time to solve a problem may be measured continuously. If the interval h is taken smaller and smaller, then the probability of solution within interval h, $\varphi(h)$, also decreases but the ratio $\psi(h)/h$ will (in any application of the present theory) approach a limit. Let

$$\lambda = \lim_{h \to 0} \left(\frac{\varphi(h)}{h} \right)$$

and it can be shown by standard methods that for a one-stage process, the density function describing the distribution of problem solving times is

$$f(t; \lambda) = \lambda e^{-\lambda t} \quad [5]$$

the exponential distribution; and in the case of k equally difficult stages,

$$g(t; \lambda, k) = \frac{\lambda}{(k-1)!} e^{-\lambda t}(\lambda t)^{k-1} \quad [6]$$

the gamma distribution. The mean and variance of the gamma distribution are

$$\mathrm{E}(t) = k/\lambda = \mu \quad [7]$$

$$\mathrm{Var}(t) = k/\lambda^2 = \sigma^2 \quad [8]$$

The exponential and gamma distributions are conceptually just like the geometric and Pascal distributions given above; in fact they are limits taken by going to continuous time (intervals of duration zero).

Estimates of Parameters

A gamma distribution with parameter k (k stages) is the distribution of the sum of k exponentially distributed, independent, random variables. An exponentially distributed variable has mean equal to standard deviation, $\mu = \sigma = 1/\lambda$, from Equation 7 and 8. When one adds independent random variables, the mean of the sum is the sum of the means, and the variance of the sum is the sum of the variances. With more stages, accidental fluctuations tend to cancel out somewhat, so that the standard deviation becomes smaller relative to the mean, the more stages are involved. With increasing k, the gamma distribution becomes more nearly symmetrical, and narrower. This property can be used to estimate the parameter k. Note that from Equations 7 and 8,

$$\mu^2 = \frac{k^2}{\lambda^2}$$

and

$$\sigma^2 = \frac{k}{\lambda^2}$$

so that

$$\mu^2/\sigma^2 = k \quad [9]$$

Given large samples of observations, it is possible to substitute the obtained mean and variance, $\bar{t}$ and s_t^2 for the parameters, obtaining the estimation equation

$$\hat{k} = (\bar{t})^2/s_t^2 \quad [10]$$

The authors have used this estimate, an example of the method of moments, instead of possible maximum likelihood or minimum chi square estimates. The method-of-moments estimate seems relatively insensitive to imperfections in the experimental procedure, and its general significance remains sensible even if the actual stages of the problem are not exactly equal in difficulty. If the stages differ in difficulty, then $\hat{k}$ can be considered as follows; the actual problem is represented by a model problem with equally difficult and independent stages, and k is the number of stages in the model. The interpretation is like one common in information theory; if a certain system transmits just 3 bits, it can be represented by a system which perfectly discriminates exactly 8 equally probable signals. The value of 3 bits may actually be derived from a system which makes imperfect discriminations of 24 categories which are not equally probable, but the model gives a general impression of the behavior of the system. Similarly, our use of a gamma distribution gives a general impression of the number of stages in a problem, of its "complexity," without necessarily being correct about the detailed structure of the problem.

Application

Davis (1961) gave three problems to 178 undergraduates, and timed each subject on each problem. The order of the three problems was changed from subject to subject. A sizable fraction of the subjects arrived at wrong answers on some problems, and the wrong answers will be set aside; even if there are exactly k stages to solution of the problem, there is no assurance that it also takes k stages to arrive at a wrong answer.

On the basis of subsidiary measurements, a relatively constant "reading time" could be subtracted from the total time consumed on each problem. When this was done, means and standard deviations of net solution time were calculated on each of the three problems.

The first (Rope) problem described a prisoner attempting to escape from a tower, who found in his cell a rope which was half long enough to permit him to reach the ground safely. He divided the rope in half and tied the two parts together and escaped. How could this have been? The answer is that the rope is divided lengthwise. This problem was taken to be a "one-stage" problem in that the solution seems to depend on a single idea.

The net mean time was 131.3 seconds, the standard deviation was 115.5, whence the estimated number of stages is $(131.3)^2/(115.5)^2 = 1.3$ stages, not far from the one stage predicted. In fact, it is possible that the time taken to write the solution might produce the modest discrepancy.

The second (Word Tangle) problem was a more complex one: "If the puzzle you solved before you solved this one was harder than the puzzle you solved after you solved the puzzle you solved before you solved this one, was the puzzle you solved before you solved this one harder than this one?" Subjects were to answer yes or no and why. The answer (yes) requires canceling of several mutually defeating phrases. Net mean time was 264.8 seconds, with a standard deviation of 154.1, so that the estimate of k is 3.0 stages.

The third (Gold Dust) problem was a more complex version of the standard water-jug problem. Given containers of 163, 14, 25, and 11 ounces, obtain exactly 77. Solutions are to subtract, from 163, either $2(25) + 14 + 2(11)$, or $25 + 2(14) + 3(11)$, i.e., either five or six manipulations. The net mean time was 373.0 seconds, standard deviation 166.7, whence the estimated number of stages is $\hat{k} = 5.0$.

The estimates of number of stages of of these problems seem, to the present writers, quite in accord with our intuitive judgments of number of stages. When a group of students, after having solved the problems, were asked to estimate the number of stages in the three problems, their mean judgments were 2.8, 4.0, and 6.3, respectively. These judgments are consistently higher than those obtained from the distributions of solution times, but the correlation is remarkable and indicates that our judgments as to the relative number of stages in the three problems agree in substance with judgments of student solvers.

In addition, it was found that the gamma distributions (Equation 6) give satisfactory fits to the obtained distributions of solution times on the three problems. The obtained cumulative distributions of time to correct solution of the three problems are shown in Figure 1, along with fitted gamma distributions. An investigation of overall goodness of fit of the theoretical gamma distribution was made separately for each problem, using the Kolgomorov-Smirnov one-sample test. The maximum discrepancies between theoretical and observed distributions were all well within the .05 acceptance region for all three problems.

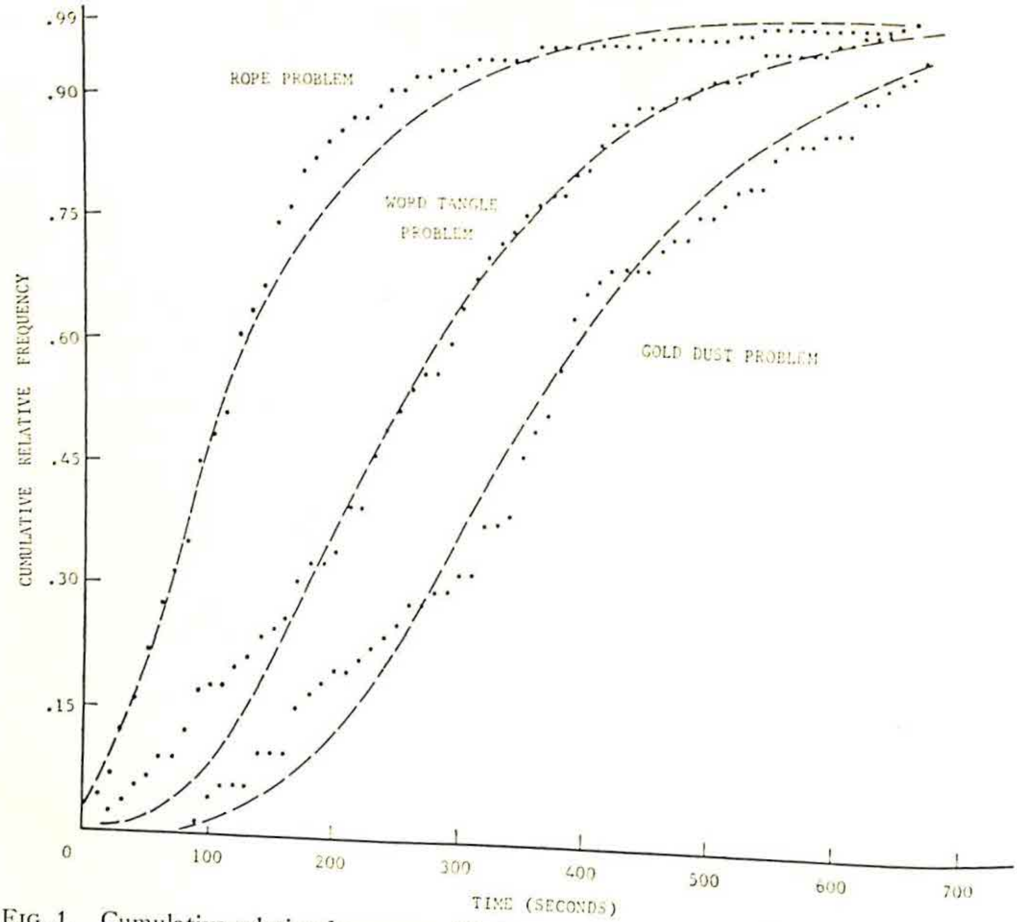


FIG. 1. Cumulative relative frequency of individual solvers and fitted gamma distributions with k = 1.3, 3.0, and 5.0 for Rope, Word Tangle, and Gold Dust Problems, respectively.

Meaning of the Model for Distribution of Solution Times

The model for the distribution of solution times assumes that the times subjects take to solve a problem vary widely because solving a problem involves having only one or a few ideas, in sequence; and the time to an idea is taken to be a random variable. This seems a natural consequence of thinking of such a problem solving process as "insightful." The detailed model includes other, less attractive, simplifying assumptions. One is that the stages of a problem are all equally difficult. Another is that the same parameter λ applies to all members of the population tested. These assumptions are probably wrong in detail, though small deviations may not be easily detectable from the kind of data we have used. It is emphasized that they are only simplifying assumptions which are made, in this model, because we were not able to separate and measure the stages of solution of the problems (because we did not, originally, know what they might be, and we are not sure now), and we did not have good individual differences predictors for our subjects.

POOLING OF ACCOMPLISHMENTS IN GROUPS

In the usual experiment on problem solving by groups, groups are better than individuals. If "Eureka" problems are used, the group is likely to solve successfully as soon as the first group member arrives at the correct solution. If it is assumed that group members are just as effective in groups as they would be working alone, then it is possible to predict group from individual performance, provided that individuals and members of group are sampled randomly from the same population of subjects, and groups are formed randomly.

The one model available for such predictions is by Lorge and Solomon (1955) who divide the population into "solvers" and "nonsolvers," use tests of individuals to estimate P_I (the proportion of solvers among individuals) and then predict that the probability of finding at least one solver in a group of size r will be $P_r = 1 - (1 - P_I)^r$. Using this theory it is possible to transform the cumulative distribution of individual solvers, which gives P_I as a function of t, into a cumulative distribution for groups of size r, though Lorge and Solomon do not do so.

Lorge and Solomon also note that if a problem has more than one part or stage, it is possible that a group will solve it though no individual in the group can solve it. Different group

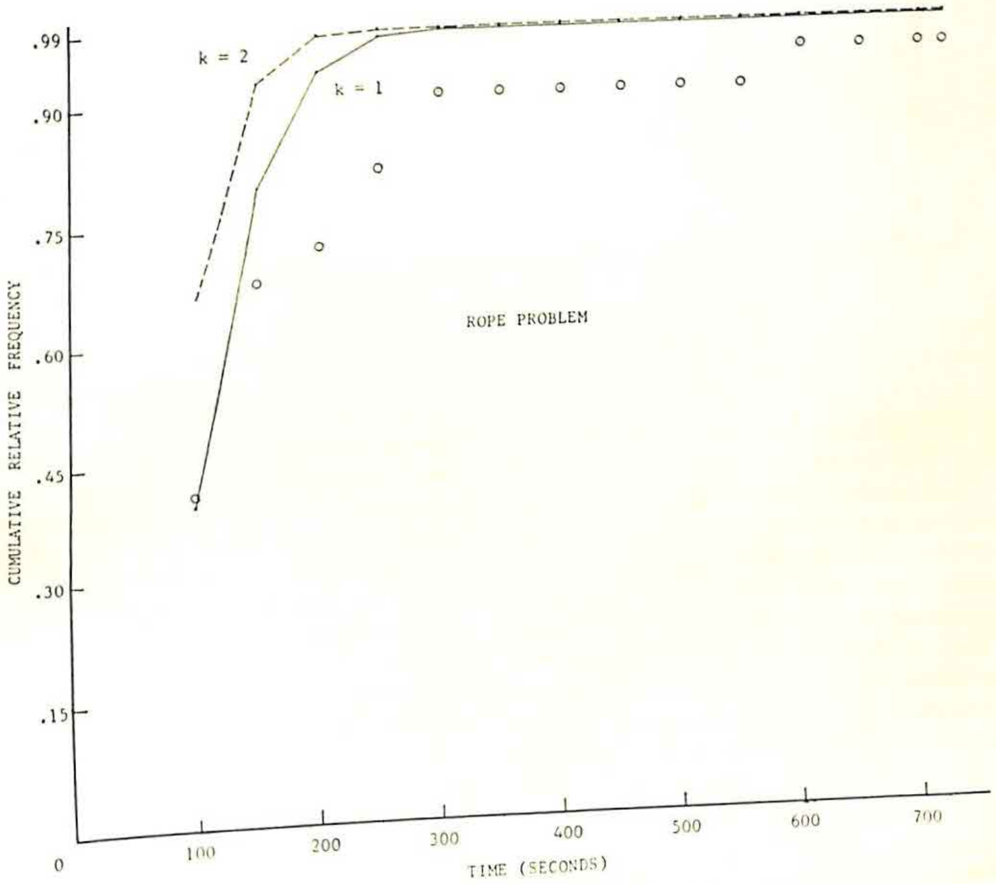


FIG. 2, Part 1.

FIG. 2. The solid and broken lines represent cumulative relative frequency at 50-second intervals, predicted by Lorge-Solomon Model ($k = 1, k = 2$, respectively) for groups of Size 4. (Observed group performance, for the same intervals, is given by solid points.)

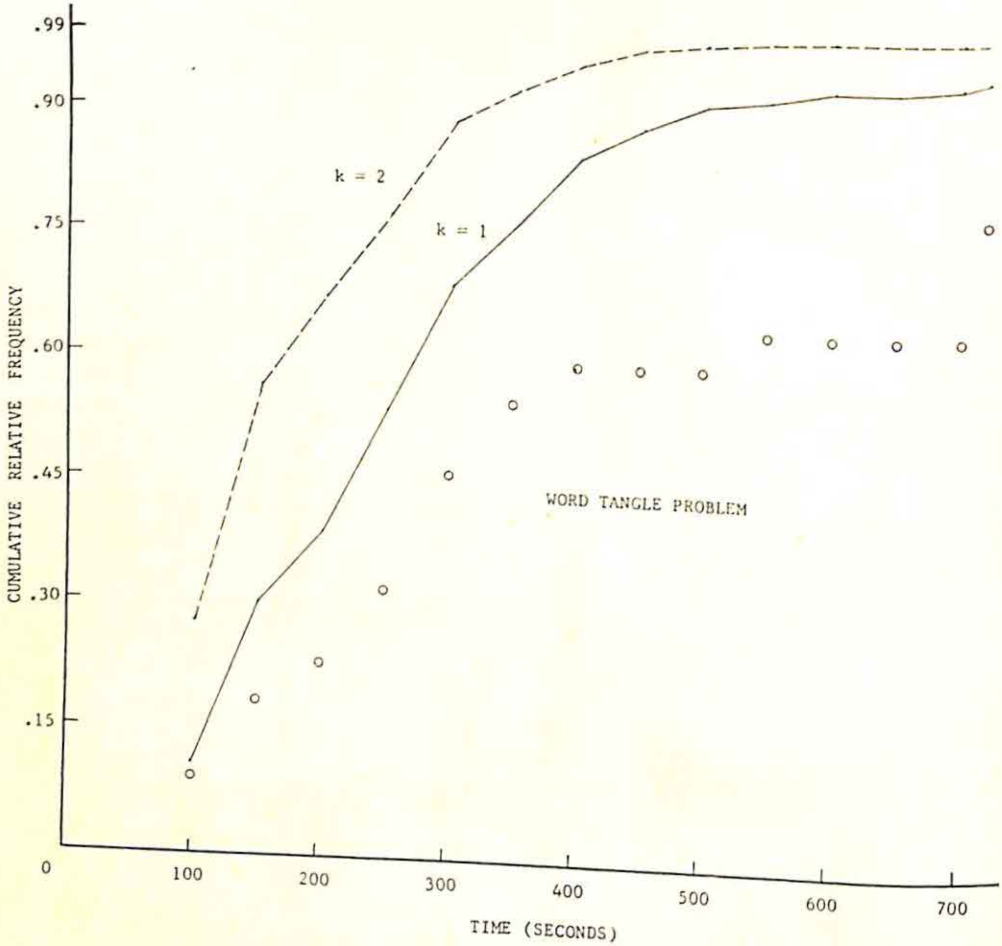


FIG. 2, Part 2.

members may be able to solve different parts, and then pool their abilities to obtain an emergent solution. If the abilities to solve the various parts are equally frequent in the population and are independent, this gives a decided extra advantage to groups. With k equally probable stages the probability of a group solution with r people is:

$$P_r = [1 - (1 - P_I^{1/k})^r]^k \quad [11]$$

Equation 11 also can be used to generate a distribution of group solution times given a distribution of individual solving times, for any given value of k.

The use of the waiting-time model for the distribution of solution-times yields an estimate of k for a problem, from the distribution of individual solution times. That estimate can be inserted for k in Equation 11, and group performance can be predicted from individual performance. Davis (1961) performed the appropriate experiment. In addition to the 178 individuals, he had 22 four-person groups work on his three problems. Recruitment of subjects, working conditions, instructions, etc., were closely comparable for individuals and groups. When Equation 11 is used to predict group from individual solution times, the results are very inaccurate. Figure 2 shows the predicted group

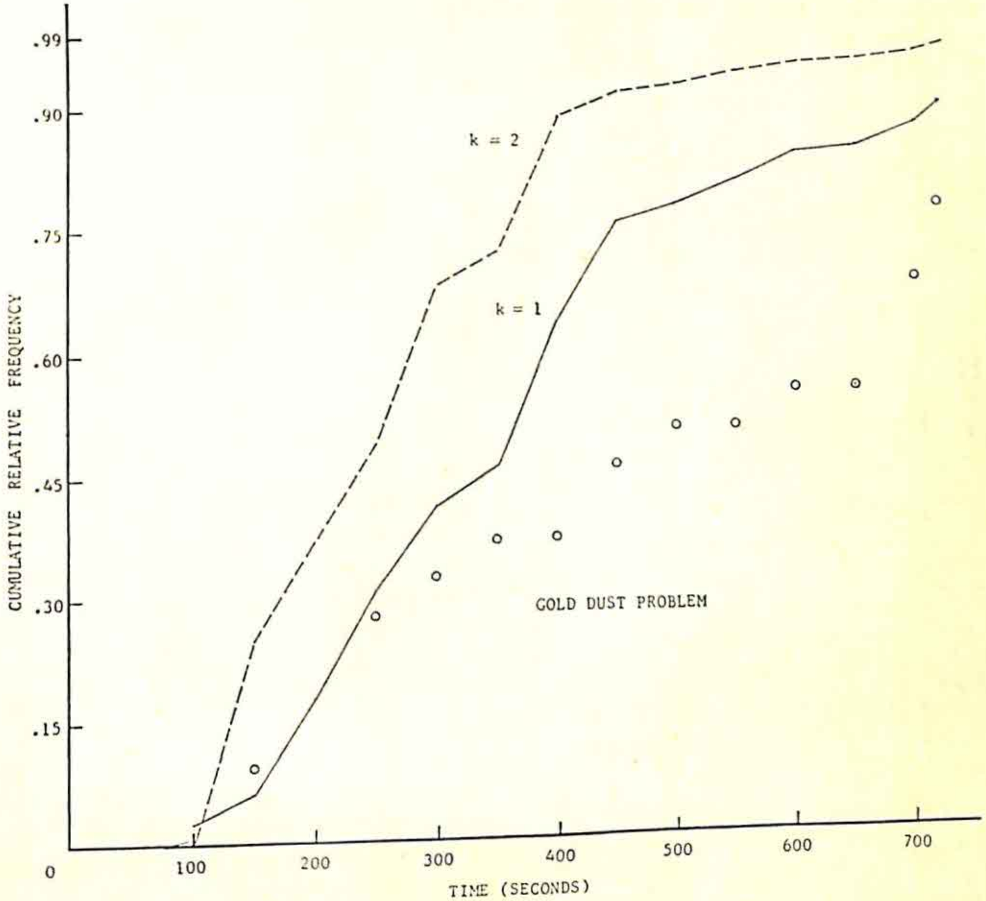


FIG. 2, Part 3.

performance (from Equation 11) and the obtained distributions on the Rope, Word Tangle, and Gold Dust problems for $k = 1$ and $k = 2$. The actual estimates of k were 1.3, 3.0, and 5.0, which would obviously give extremely inaccurate estimates.

However, deeper consideration shows that the Lorge-Solomon model is not the proper one to be used with the waiting-time model developed. Lorge and Solomon begin by dividing the population into solvers and nonsolvers. To obtain more complete predictions, we extended the idea to "solvers by time t" and "nonsolvers at time t," which seems a legitimate extension of the ideas of Lorge and Solomon. However, we earlier had generated the distributions of solution time from the assumption that solution was a random event, and all subjects had equal probability of solution in a given interval. The Lorge-Solomon model takes the individual performance as a fixed fact which identifies solvers and nonsolvers, assumed to be individuals with different abilities. The random waiting-time model leads to a different, combination-of-*contributions* model.

Combination of Contributions

Consider a group of subjects none of whom misinterprets the problem or arrives at a wrong solution. Suppose that for each subject the problem involves the same k stages each of which has an exponential distribution of time-to-completion of $\lambda e^{-\lambda t}$. With r subjects in a group, it can be shown that the distribution of group performance is given by the gamma distribution,

$$g(t; r\lambda, k) = \frac{r\lambda}{(k-1)!} e^{-r\lambda t}(r\lambda t)^{k-1} \quad [12]$$

which is precisely the same gamma distribution derived for individual performance except that the rate (λ) is increased by a factor of r, the size of the group. Part of the proof is given. Within a short interval h, given that no one has solved a certain stage, the probability that some one in the group will solve is $1 - (1 - \Psi)^r$. As h is taken small, Ψ becomes small, in which case

$$1 - (1 - \Psi)^r = 1 - 1 + r\Psi - \binom{r}{2}\Psi^2 + \cdots \pm \binom{r}{r}\Psi^r$$

approaches $r\Psi$. The reason is that all terms involving Ψ^2, Ψ^3, etc., become relatively negligible when Ψ becomes small. Hence, in the limit, the distribution of time to a given stage approaches $r\lambda$.

This calculation provides a prediction of group problem solving if no subjects arrive at wrong solutions. Preliminary results of an experiment by R. Hoppe at Michigan State University[4] give some support for this model. Hoppe used Rokeach's (1960) "Doodlebug" problem and tested individuals, and groups of Size 2 and 3. In Hoppe's study the subjects could ask the experimenter questions and were, in the main, redirected when they had made a serious mistake, so that all subjects and groups either solved the problem correctly or ran out of time. The data were in agreement with the hypothesis that all time distributions were of the same shape (it is not possible to make a good test of the gamma distribution since a substantial proportion of subjects failed to finish) and that the rate of solution is roughly proportional to the number of people in the group.

In Davis' (1961) experiment the subjects received no redirection from the experimenter, and about half of the individuals arrived at wrong answers on the two harder problems (Word Tangle and Gold Dust). The effect of wrong answers in the group situation is not directly known, since observations did not enable the experimenter to tell which members of groups had solved the problem. However, two plausible hypotheses can be formulated, and they lead to two quite distinct predictions regarding group performance.

Hierarchical model. It is possible that subjects going "off the track," referred to here as nonsolvers, are nonfunctional in their groups. If so, a group with two solvers and two nonsolvers will solve the problem just as rapidly, on the average, as a group of two solvers alone. (It is assumed that in Eureka problems, groups will not agree on a wrong solution.) Nonsolvers would be nonfunctional if, for example, the solvers suppress nonsolvers and form an intellectual hierarchy within the group.

If a group of size r has A solvers and $B = r - A$ subjects who are nonsolvers, then by the Hierarchical model the distribution of time to solution should be a gamma distribution, $g(t; A\lambda, k)$. If a proportion a of individual workers solve the problem, then the probability that a group of size r has exactly A members who will solve is

---

[4] Ronald Hoppe, personal communication, 1961.

$$P(A) = \binom{r}{A} a^A (1 - a)^{r-A}, \quad [13]$$

a binomial distribution.

The distribution of solution times of a collection of groups will then be described by the density function,

$$f_H(t) = \sum_{A=0}^{r} P(A) g(t; A\lambda, k) \quad [14]$$

if the hierarchical model is correct.

Equalitarian model. Another possibility is that those subjects who go off the track, the nonsolvers, continue to take full part in group discussion though they do not arrive at solutions. This idea leads to the hypothesis that the nonsolvers would consume their share of the group's time, slowing up the other members of the group. This would mean that the probability that a solver in the group solves in unit time is reduced from λ to $[A/(A + B)]\lambda$, when there are A solvers and B nonsolvers present. The distribution of solution times for such a group is, by the argument leading to Equation 12,

$$g\left(t; A\,\frac{A}{A + B}\,\lambda, k\right) = g\left(t; \frac{A^2}{A + B}\,\lambda, k\right) \quad [15]$$

Given that the probability of an individual's solving is a, the distribution of solution times of a collection of groups will be described by the density function,

$$f_E(t) = \sum_{A=0}^{r} \left[P(A) g\left(t; \frac{A^2}{A+B}\,\lambda, k\right)\right] \quad [16]$$

if the equalitarian model is correct.

The evidence of Davis' (1961) experiment is that the hierarchical model is seriously in error, and the equalitarian model yields predictions which are quite close to the data, not seriously in error. The Kolgomorov-Smirnov one-sample test leads to rejection of the distribution calculated from the Hierarchical Model, for the word tangle and gold dust problems, at the .05 level of significance. The distribution derived from the Equalitarian Model is accepted at the .05 level for all three problems. Of course, with only 22 groups, the statistical acceptance is not decisive. Predictions from these two combination-of-contributions models, along with the distributions of group performance, are shown in Figure 3.

Subsidiary evidence in the form of sociometric choices and communication frequencies (Davis, 1961) was collected, and the results agree in substance with the assumption of an equalitarian group structure.

DISCUSSION

Structure of Problems

The estimates of number of stages of the Rope, Word Tangle, and Gold Dust problems would be very impressive evidence, except that we have no strong independent measurement of the number of stages in such problems. The intentions and judgments are suggestive but not decisive. The number of stages in a problem is the number of distinct ideas required to obtain the solution. The number of stages is the number of ideas which must be conjoined to obtain the solution—the subject must have Idea 1 *and* Idea 2 *and* . . . , etc. If there were alternative answers, the subject could have Idea 1 *or* Idea 2 *or* . . . , and these would not be stages. This analysis makes it possible to compare stage-estimates from the data with the logical structure of the problems; and

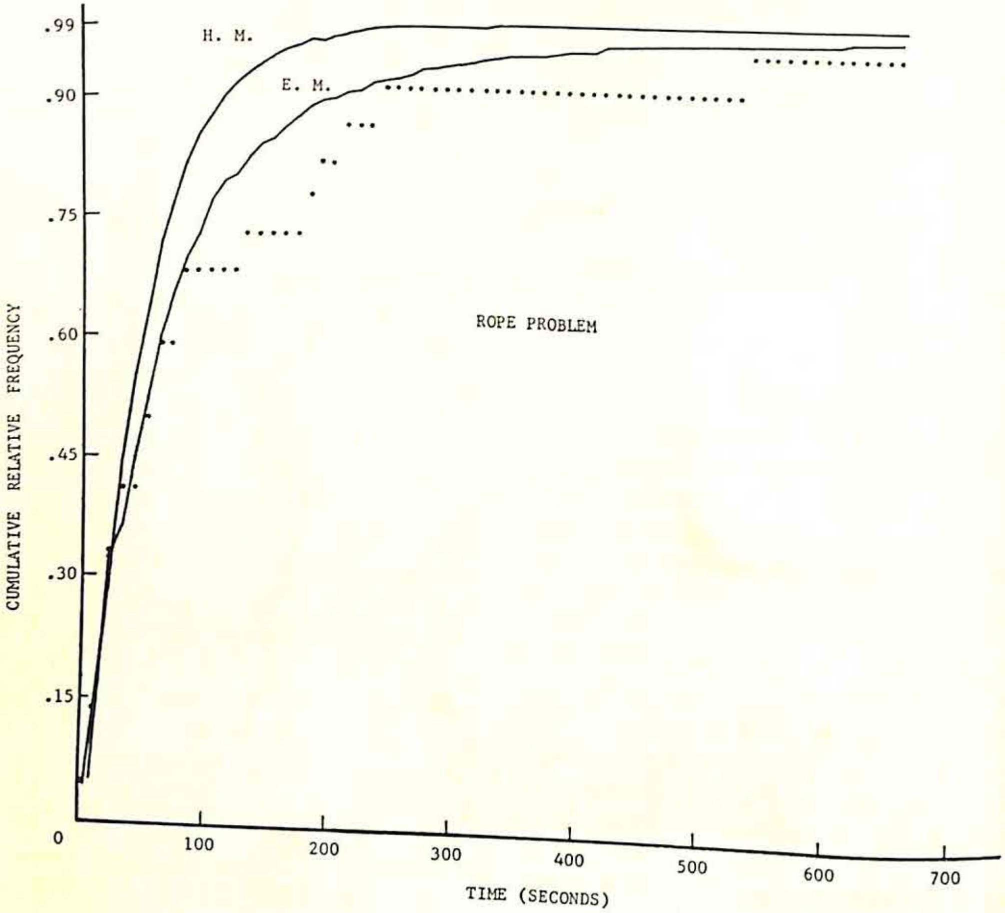


FIG. 3, Part 1.

FIG. 3. The solid lines represent cumulative relative frequency predicted by the Hierarchial and Equalitarian Models for groups of Size 4. (Observed group performance is given by solid points.)

the agreement is good. Unfortunately, we are not sure how to identify "ideas" in the sense used above, hence the logical analysis is not sure of its elements.

What is needed, obviously, is an experimental separation of stages in problem solving. Hoppe's work, mentioned above, attempts such an analysis. A difficulty is that one can hardly separate the stages of a problem, experimentally, unless he knows what those stages are. Verbal protocols of problem solving process are useful for identifying parts, but they present difficulties of interpretation. The writers do not foresee any easy and immediate answer to the question of problem structure.

Stages versus parts. In the special model above it was assumed that the stages are to be accomplished in sequence. If there are simply k parts which must be accomplished, in any order, then the distribution of times may be changed. If the subject is thought to work on only one stage at a time, solving the stages in some irregular order, then the gamma distribution model applies. However, if ideas for different parts of the problem can arise freely and in-

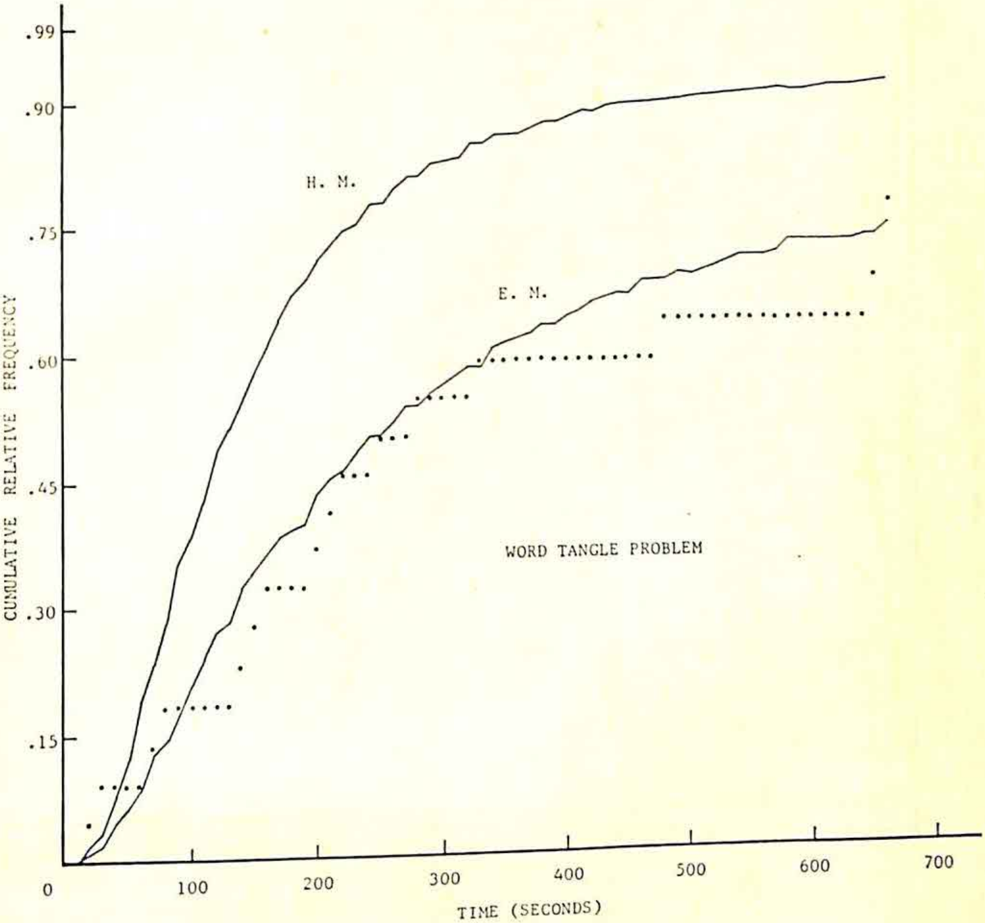


FIG. 3, Part 2.

dependently, a new distribution arises. Solution time of any part will be exponentially distributed, but until the subject has solved the first part there are k parts which may be solved. Hence the rate until the first part-solution is $k\lambda$. After one part is completed, there are $k-1$ parts left and the rate until the second part is $(k-1)\lambda$. Finally, when there is only one part left, the rate is λ. These "rates" can be defined as above if the distribution of times until the next stage is assumed to be exponential.[5]

The resulting distribution is rather complicated, but its mean and variance can be developed easily. Since the mean of each of the waiting times is independent of the others by hypothesis, the total mean is the sum of k waiting times with means $1/\lambda$, $1/(2\lambda), \cdots, 1/(k\lambda)$. Hence the mean total time to solution, $E(T)$, is

$$\mathrm{E}(T) = (1/\lambda) \times (1+1/2+1/3+\cdots+1/k) \quad [17]$$

By a similar argument, the variance of total times is

$$\mathrm{Var}(T) = (1/\lambda^2) \times [1+(1/2)^2+(1/3)^2+ \cdots +(1/k)^2] \quad [18]$$

---

[5] The writers are indebted to Herman Rubin for the formulation and solution of this mathematical problem.

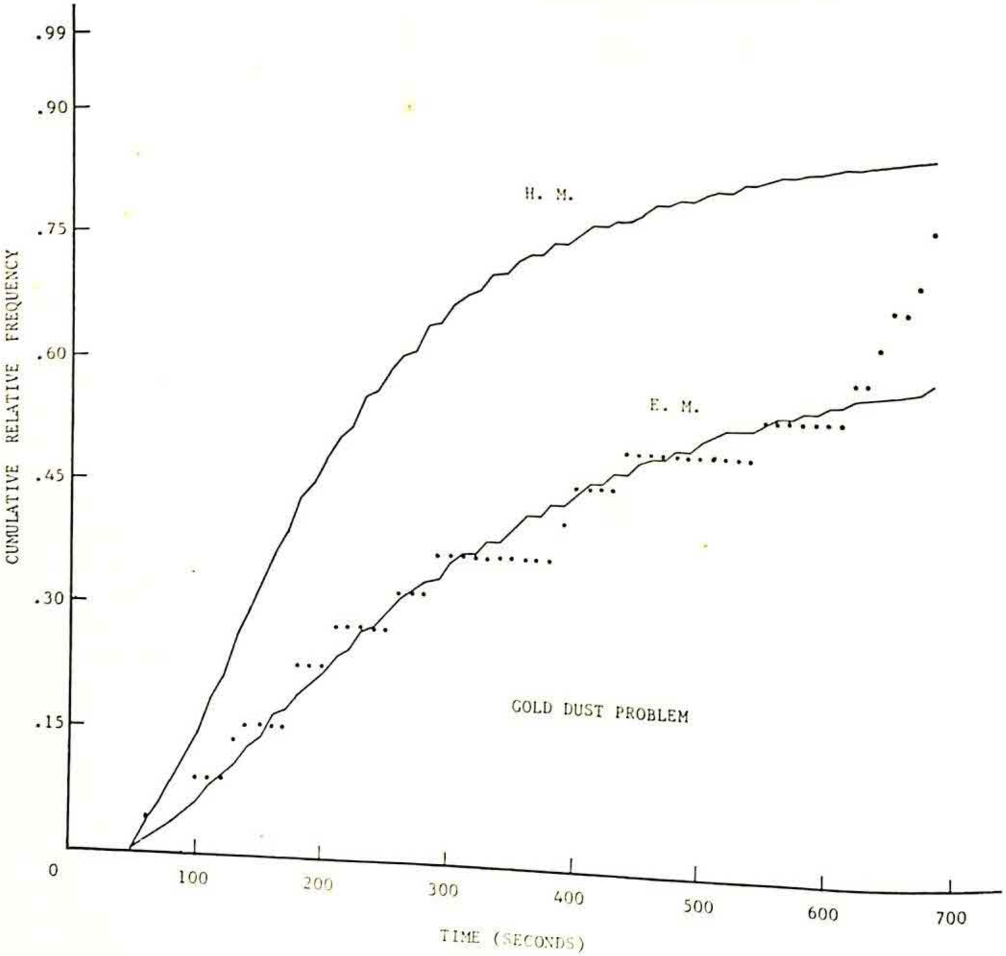


FIG. 3, Part 3.

TABLE 1

SOME STATISTICS OF PROCESSES WITH PARTS
(Which May be Solved Concurrently)
AND PROCESSES WITH STAGES
(Which are Solved Successively)

Number of parts or stages	$\mu\lambda$		$\sigma\lambda$		μ^2/σ^2	
	Parts	Stages	Parts	Stages	Parts	Stagse
1	1.00	1.00	1.000	1.000	1.000	1.00
2	1.50	2.00	1.108	1.414	1.800	2.00
3	1.83	3.00	1.117	1.731	2.469	3.00
4	2.08	4.00	1.119	2.000	3.049	4.00
5	2.28	5.00	1.210	2.238	3.562	5.00
6	2.45	6.00	1.221	2.452	4.025	6.00
7	2.59	7.00	1.229	2.642	4.447	7.00
8	2.72	8.00	1.236	2.828	4.836	8.00
9	2.83	9.00	1.241	3.000	5.198	9.00
10	2.93	10.00	1.245	3.161	5.536	10.00

Table 1 shows how mean and standard deviation, and also the value of μ^2/σ^2, progress with more and more parts. Comparable values for "stages," i.e., parts which must be solved in succession, are shown for contrast.

Possible effects of wrong solutions. The Equalitarian model for predicting group from individual performance was chosen because it seemed to take proper account of the fact that some subjects may arrive at wrong but self-satisfying answers. However, wrong solutions may distort the distribution of times to correct solution. In estimating the number of stages to solution and working out the waiting-time model, subjects who arrived at wrong solutions were eliminated from the data. This seems reasonable, for the number of stages to a wrong solution may be different from the number of stages to a correct solution. However, this conditional distribution of times (times to solution given that a correct solution is reached) may be misleading. The difficulty is that a subject who happens to take a long time to get the correct answer has an increased vulnerability to error, whence long times to correct solution may be systematically less frequent when errors occur than they would be if the whole distribution were known.

The difficulties can be studied by investigating the properties of random systems with more than one outcome. The most general model available to the present writers is that proposed by LaBerge (1960)[6] which permits any of several responses which involve k_1, k_2 . . . stages respectively. LaBerge derives the distribution of times to completion of the k_1 stages of Outcome 1 provided that Outcome 1 occurs.

One special model may be called the "misreading" model, in which it is supposed that the subject may, at the very beginning, make a choice which leads to a wrong solution. The subject either enters into a process which leads to success (with the gamma distribution of Equation 8), or onto some other which leads to failure. The conditional distribution of times to correct solution, given that a correct solution and not an error is attained, is the original gamma distribution.

A third possibility is that the subject may at any time before solution make a decision which leads to a wrong answer. Consider that this wrong decision is a unitary event which has a constant probability of happening at any time, given it has not already happened, up until the problem is solved. Then the accidental event will have an exponential distribution. The density (probability per time interval) of solution of the problem will be the product of the original gamma density, $g(t; \lambda, k)$ and the probability that the accidental event has not occurred. If the constant probability-per-unit-time or rate of the accidental event is α, the probability that it has not occurred by time t (provided the problem is not solved before t) is $e^{-\alpha t}$. The conditional density of solution at time t, given that a correct solution is attained, is

$$f(t; \lambda, \alpha, k)$$

$$= \frac{g(t; \lambda, k)e^{-\alpha t}}{\int_0^\infty g(t; \lambda, k)e^{-\alpha t}dt} \qquad [19]$$

$$= \frac{\frac{\lambda}{(k-1)!}\, e^{-(\lambda+\alpha)t}(\lambda t)^{k-1}}{\int_0^\infty \frac{\lambda}{(\mathrm{k}-1!)}\, e^{-(\lambda+\alpha)t}(\lambda t)^{k-1}dt} \qquad [20]$$

The numerator is simply

$$g(t; \lambda+\alpha, k)\left(\frac{\lambda}{\lambda+\alpha}\right)^k \qquad [21]$$

and the denominator is

$$\left(\frac{\lambda}{\lambda+\alpha}\right)^k \int_0^\infty g(t; \lambda+\alpha, k)dt$$

$$= \left(\frac{\lambda}{\lambda+\alpha}\right)^k \qquad [22]$$

The step from Equation 20 to Equation 21 is accomplished by noting that $g(t; \lambda+\alpha, k)$

$$= \frac{\lambda+\alpha}{(k-1)!}\, e^{-(\lambda+\alpha)t}[(\lambda+\alpha)t]^{k-1}$$

$$= (\lambda+\alpha)^k e^{-(\lambda+\alpha)t} t^{k-1}$$

---

[6] LaBerge, D. A statistical theory of discrimination behavior. (Prepublication draft, December 1960.)

If this is multiplied by $[\lambda/(\lambda+\alpha)]^k$ one obtains the numerator of Equation 20.

Hence the conditional distribution of solution times is

$$f(t;\lambda,\alpha,k) = g(t;\lambda+\alpha,k) \quad [23]$$

a gamma distribution with the original number of stages, k, but with rate increased to $\lambda + \alpha$. The probability of a correct solution is $[\lambda/(\lambda+\alpha)]^k$.

In summary, three effects of wrong solutions were mentioned; they might distort the distribution from a gamma to some other distribution (if the subject works concurrently on stages of several different processes leading to several answers), it might leave the conditional distribution of solution times unchanged (if errors arise only through an initial wrong decision such as misreading the problem), or it might leave the conditional distribution a gamma, still with k stages, but increase the rate of solutions, (if entry onto an erroneous path is thought of as a unitary random event.) Of these, the third seems most in accord with the problem solving situation in Davis' (1961) experiment.

Unhappily, we do not have the data needed to decide which of these possibilities, if any, occurred in the Davis (1961) experiment. As we understand it, the LaBerge formulation could be fit to the data exactly only if we knew what all the wrong solutions were and could estimate the number of stages and the rate parameters of each. Davis' experiment was not designed to facilitate such an investigation, and the necessary analysis is not feasible. Since LaBerge's model involves parameters we cannot estimate, it cannot be compared with the other models validly. The misreading model and the accidental event model do not require analysis of the inner structure of erroneous solutions, hence can be applied to the present data. Since both lead to gamma distributions of correct solution, both agree with Davis' results. Only a detailed qualitative analysis of protocols, or a technique for eliminating wrong solutions without otherwise disturbing performance, would indicate which model is correct. In fact, some subjects might misread and others err through an accidental random event, and the conditional distribution of correct solution time would still be a gamma distribution.

This theoretical result increases the plausibility of the gamma distribution as an hypothesis by showing that it might, under reasonable conditions, characterize data from problems with quite different frequencies of wrong solution.

OTHER MODELS FOR OTHER KINDS OF GROUPS

The predictions of group performance in Davis' experiment were for ad hoc groups of equivalent members, with no discernable social inequalities. Furthermore, the nature of the problems was such that ideas were pooled with little controversy over the correctness of a correct proposal. In a "team" it sometimes happens that all members must solve their parts of the problem before the group succeeds. Then the probability of group solution is the joint probability that all members have completed their tasks by time t. If the individual assignments are unitary problems with the same difficulty, and all members of the group can work concurrently (as in servicing an automobile) then the model leading to Equation 17 and 18 should describe the situation. If one stage of the process must wait upon completion of another stage, then combining the individual contributions can lead to a gamma distribution. If solution of the problem is mainly a matter of decision rather than invention, then the probability of decision by time *t* may be the probability that all group members accept the same idea at that time. Changes of opinion, persuasion, and the invention of compromises are the events of primary interest, and the main results of the present paper do not apply.

These remarks emphasize that the mathematical theory used in this paper is not a general theory of group problem solving, but is a careful and precise formulation of some hypotheses about one particular kind of group problem solving. The hypotheses depend both upon the constitution of the groups and the nature of the problems.

RELATION TO OTHER AREAS OF PSYCHOLOGY

The mathematical forms discussed in this paper have applications elsewhere in psychology. The notion that an individual stage of problem solving will have an exponential distribution of solution time, is essentially the same as the hypothesis that a unitary learning event will occur on a geometrically distributed trial. Both arise from the statement that the event has constant probability of occurring, until it does occur. In learning theory, this is the hypothesis of "all-or-nothing" learning (Bower, 1962; Estes, 1960; Restle, 1961b, 1962). A truly unitary process, indivisible and independent of other processes, should have a constant probability of occurrence. If an event is not unitary but a compound, then every effort should be made to tease apart the stages involved in it. Hence the geometric and exponential distributions (for trials and continuous time, respectively) give a theory for any event which can properly be left unanalyzed.

The only compounds which do not require detailed analysis into parts are those made up of equivalent parts. A beam of light can be thought of as a compound of unitary emissions of photons, but the only interesting question about such a beam is the number of photons, since separate photons are all alike. Thus, an ideal compound event must be one made up of equivalent, identically distributed parts. The gamma distribution describes compounds of the type used in this paper. Previous applications of gamma distributions include Bush and Mosteller (1955, Ch. 14) on the time rats take to traverse a runway, and Restle (1961a) on simple reaction time.

Similarly, the simplest kind of model for comparing group with individual performance is one which assumes that people in a group are, in all ways possible, just like people working alone. In a nonevaluative sense one can say that the Lorge-Solomon models, and the variations given in the present paper, are all versions of this ideal hypothesis.

Any deviations between model and data suggest the action of factors more complex (and perhaps more interesting) than the very simplest possibilities discussed here. For example, since Davis' groups did not reach the level predicted by Lorge-Solomon models, it became possible to introduce the hypothesis that individuals who have entered on a wrong interpretation of the problem take up their share of group time and interfere with the progress of others. Discrepancy from one model became a measurement of a previously unmeasured factor.

SUMMARY

A waiting-time model was developed to give a distribution of times to solution, for word puzzles. The distribution depends upon the number of stages involved in solving the problem. Experimental results showed that the theoretical distribution (gamma distribution) gives a good fit to the data, and the estimated number of stages agrees well with intuition and judgments. A theory of "pooling of contributions" was developed to describe the superiority of groups over individuals in solving word puzzles. Experimental results suggested that subjects who have misunderstood the problem and will arrive at wrong answers, nevertheless consume their share of group time. With this addition, the pooling of contributions theory accurately predicts group problem solving data. Some alternative models, resulting from different problem structures and different kinds of groups, were discussed, and the waiting-time theory was shown to be related to the concept of "all-or-nothing" learning.

REFERENCES

BOWER, G. H. An association model for response and training variables in paired associate learning. *Psychol. Rev.*, 1962, **69**, 34–53.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

DAVIS, J. H. Models for the classification of problems and the prediction of group problem-solving from individual results. Unpublished doctoral dissertation, Michigan State University, 1961.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

FAUST, W. L. Group versus individual problem solving. *J. abnorm. soc. Psychol.*, 1959, **59**, 68–72.

FELLER, W. *Introduction to probability theory and its applications.* New York: Wiley, 1957.

HUSBAND, R. W. Cooperative versus solitary problem solution. *J. soc. Psychol.*, 1940, **11**, 405–409.

LORGE, I., FOX, D., DAVITZ, J., & BRENNER, M. A survey of studies contrasting the quality of group performance and individual performance, 1920–1957. *Psychol. Bull.*, 1958, **55**, 337–372.

LORGE, I., & SOLOMON, H. Two models of group behavior in the solution of eureka-type problems. *Psychometrika*, 1955, **20**, 139–148.

MARQUART, DOROTHY I. Group problem solving. *J. soc. Psychol.*, 1955, **41**, 103–113.

RESTLE, F. *Psychology of judgment and choice.* New York: Wiley, 1961. (a)

RESTLE, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, **26**, 291–306. (b)

RESTLE, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

SHAW, MARJORIE E. Comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.*, 1932, **44**, 491–504.

TAYLOR, D. W., & FAUST, W. L. Twenty questions: Efficiency in problem solving as a function of size of group. *J. exp. Psychol.*, 1952, **44**, 360–368.

THORNDIKE, R. L. On what type of task will a group do well? *J. abnorm. soc. Psychol.*, 1938, **333**, 408–412.

WATSON, G. B. Do groups think more efficiently than individuals? *J. abnorm. soc. Psychol.*, 1928, **23**, 328–336.

(Received July 28, 1961)

Psychological Review
1962, Vol. 69, No. 6, 537–541

THEORETICAL NOTES

THE MOLECULAR BASIS OF MEMORY AND LEARNING

MICHAEL H. BRIGGS AND G. BARRIE KITTO
Victoria University of Wellington, New Zealand

In recent papers, Gaito (1961) and Hyden (1961) have suggested that nerve activity results in a rearrangement in the sequence of monomers of certain high molecular weight polymers (nucleic acids and proteins) found in brain cells, or in the synthesis of additional polymers of novel sequences, and that this rearrangement or synthesis is the molecular basis of memory. It is the purpose of this note to show that the only mechanisms whereby such rearrangements or syntheses could occur predict chemical brain changes incompatible with present experimental evidence, and, moreover, that current biochemical knowledge of nucleic acids renders the hypothesis almost certainly false. The evidence brought forward in support of polymer sequence rearrangement is then shown to be compatible with a second hypothesis that is not contrary to current biochemical views on biopolymers.

There are three main types of polymers[1] in all cells, including neurones. The deoxyribonucleic acids (DNAs) are confined to the nucleus and, in the form of nucleoproteins, are the major constituents of the genes. There are two types of ribonucleic acids (RNAs). A soluble form (messenger-RNA) is apparently synthesized in the nucleus and then transferred to the cytoplasmic ribosomes. These structures contain insoluble RNAs and are the site of protein synthesis.

The interrelations of DNA, RNA, and protein are relatively simple. Every cell contains at least one type of DNA molecule for every gene. These molecules are chemically identical to the DNAs of the parent cells. Each DNA molecule is a long, unbranched chain of nucleotides, which usually occurs chemically bonded to a second molecule in the form of a twin helix. Replication of a DNA molecule occurs only at cell division when the two components of a helix separate and each serves as a template for the synthesis of a new polynucleotide to form two new helices.

As these processes occur only at cell division, it is clear that the DNA molecules in the nucleus of a neurone are permanent structures, and there is no experimental evidence of any significant change in quantity with time.[2] The base sequence of any particular DNA molecule is determined by the base sequence of the template on which it was formed (Kornberg, 1960). It has been demonstrated that no polynucleotide synthesis occurs in the absence of a template (Kornberg, 1960).

This evidence rules out changes in the base sequence of DNAs as molecular "memory banks." There is no way in which electrical changes in the cell walls of neurones could interfere with the process of polynucleotide synthesis to give new base sequences. DNA is formed only on a template, spontaneous coupling of nucleotides does not occur. Hence, even if DNA molecules of the nucleus were fragmented by an electrical impulse, they would not recombine into polynucleotides of different base sequence.

The only known mechanism for altering the base sequence of a DNA molecule is by reaction with a mutagenic agent. These are compounds which chemically alter the purines and pyrimidines of the polynucleotide (see, for example, Freese, 1961). The effect of a mutagen is to alter the base sequence *not* by a rearrangement of the existing bases, but by the conversion of some of the bases *in situ* on the polymer molecule to other related compounds, not normally found in DNAs. It could be hypothesised that nerve impulses produce chemical mutagens *in vivo*. However, this cannot be true for no unusual bases have been found in brain DNAs, nor incidentally, in brain RNAs.

---

[1] Details of the biochemistry of DNA, RNA and their relation to protein biosynthesis can be found in numerous publications, e.g. Abrams (1961), Leslie (1961).

[2] Total brain DNA reaches a minimum in man at about age 30 years, then steadily increases. Changes in individual brain cell DNA concentrations are at present uninvestigated.

Nuclear DNAs continually serve as templates for the synthesis of messenger-RNA. The base sequence of the DNA template determines the base sequence of the messenger-RNA formed on it (Weiss & Nakamoto, 1961). Again, there is no opportunity for electrical impulses to interfere with the process to give a messenger-RNA of base sequence undetermined by its DNA template. The steric relationships necessary for synthesis on the template are absolutely specific for particular bases (Abrams, 1961).

After transfer to the ribosomes, the RNA serves as template for protein biosynthesis. The sequence of amino acids in any particular type of cell protein (including enzyme proteins) is determined directly by the base sequence of the RNA on which it forms, and, of course, indirectly by the base sequence of the DNA on which the RNA was formed (Speyer, Lengyel, Basilio, & Ochoa, 1962; Sueoka, 1961).

It is quite clear that the only way in which electrical impulses could disrupt any of these template mechanisms to give altered monomer sequences is by the formation of a mutagen. The absence of unusual bases in brain nucleic acids precludes this possibility. Moreover, even if a change in polymer base sequence was caused in a neurone by some outside influence, the effects would probably not have the results desired by Gaito (1961). Thus, unless the alteration in monomer sequence were a large one, which would certainly be disastrous for the cell due to the production of foreign protein, the only effect would be some small change in the amino acid sequence of a protein. To be of any consequence this would have to be an enzyme protein. If the change occurred at the active center, the enzyme would lose its catalytic properties and be useless to the cell. If the change occurred elsewhere in the protein molecule the chances are greatly against any significant alteration in properties (Fincham, 1960).

In summary, one may say that the hypothesis of brain polymer sequence alteration as the molecular basis of memory is unlikely because the known template mechanisms of polymer biosynthesis preclude any such changes, and even if such changes did occur they would, if large, lead to a complete disruption of cell activity, or, if small, to no significant change.

Experimental evidence in support of the polymer rearrangement hypothesis of memory has been published by several workers. This evidence is indirect and is concerned largely with the supposed role of RNAs as molecular memory-banks. It has been demonstrated that increases in total brain RNA occur with age (Cameron & Solyom, 1961), that the administration of structural analogs of purine bases affects rat maze learning (Dingman & Sporn, 1961), that treatment with ribonuclease affects the retention of conditional responses in regenerated planarians (Corning & John, 1961), and that activity produces alterations in the base ratios of RNA of brain cells (Egyhazi & Hyden, 1961; Hyden, 1959, 1961).

Considered as a whole, this evidence definitely demonstrates a role for RNA in the molecular processes basic to memory and learning. But, on logical grounds, it is necessary first to demonstrate that this role is not connected with the well-established function in protein biosynthesis, before postulating a second function involving nucleotide rearrangements that are improbable on chemical grounds. It is the purpose of the following discussion to propose that the molecular basis of memory involves changes in enzyme concentration in brain cells, and that these changes necessarily involve the observed alterations in RNA metabolism.

It is a well-known principle of genetics that all somatic cells contain the same genes. Each gene (consisting of DNA) controls the synthesis of a particular enzyme (Leslie, 1961). Hence every body-cell has the potential to make the same set of enzymes. But it is quite clear that there are large enzyme differences between cells. Some lack enzymes possessed by others, while the concentration of a particular enzyme varies widely between cells of different tissues (Dixon & Webb, 1958). It is apparent that the genes of some cells are inactive; in the sense that they are not participating in protein biosynthesis to any significant extent: while others are not functioning at their maximum rate. There must exist within cells a mechanism (or mechanisms) to initiate and control the rate of DNA, messenger-RNA, and messenger-RNA-protein syntheses. The chemical basis of these mechanisms is still obscure, but many enzymes are known to be induced by the presence of their substrates (Dixon & Webb, 1958). It is possible that the presence of substrate within a cell initiates synthesis, while the concentration of substrate determines the rate at which a particular enzyme is produced. This mechanism could be basic to the biochemical basis of memory and could operate in the following manner.

The synapses of the brain are polyfunctional, in that several axons terminate at the dendrites on any particular neurone. For this neurone to transmit an impulse it must be stimulated. The stimulus, which is chemically transmitted, can come either from a single axon, or from a combination of axons discharging together. Whether or not the impulse is transmitted depends upon whether the concentration of released transmitter substance reaches the critical level to stimulate the next neurone. Now, the amount of transmitter substance manufactured and stored by a neurone must depend upon several factors: including the activity within the cell of the biosynthetic enzyme system for the substance. If the enzymes of this system are inducible, the exchange of metabolites and demands for transmitter substance on repeated stimulation will increase the neurone's ability to manufacture the substance. Hence on the first few stimulations a neurone may not release sufficient transmitter substance to produce a postsynaptic response, but the amount of substance released on subsequent occasions will be sufficient.

If this hypothesis is incorporated with the concept of polyfunctional synapses and interconnecting neurone pathways in the brain, it becomes clear that an increased amount of transmitter substance released by any axon following repeated stimulation could be the major factor in establishing and maintaining the particular nervous mechanisms of memory and learned behavior.

Experimental investigation of the hypothesis could be attempted by a determination of whether the enzyme systems for the biosynthesis of nerve transmitter substances are inducible, and whether there is any increase in the amount of transmitter substance released by an axon following repeated stimulation. The recent reports of variations in cholinesterase activity with age, strain, individuals, and after learning (Aprison & Himwich, 1954; Bennett, Krech, Rosenzweig, Karlsson, Dye, & Ohlander, 1958; Bennett, Rosenzweig, Krech, Karlsson, Dye, & Ohlander, 1958; Krech, Rosenzweig, & Bennett, 1956, 1959; Krech, Rosenzweig, Bennett, & Longueil, 1959) in the brain of rats, could well be changes induced by increased amounts of the transmitter substance acetylcholine, which is being formed by neurones in larger quantities due to increased choline acetylase activity as suggested by the above hypothesis.

It will be seen that the hypothesis suggested here, that the molecular basis of memory and learning is basically induced changes in the enzyme systems involved in the formation and degradation of nerve transmitter substances, unlike the hypothesis of nucleotide sequence rearrangement, makes no proposal contrary to current biochemical views. Moreover, increased concentration of specific enzymes necessarily implies increased activity of the RNA-dependent biosynthetic systems for these enzymes. In fact, certain definite changes in RNA metabolism are predicted by the hypothesis. These are outlined below.

Increases in enzyme concentrations imply increases in RNA. Considering the central nervous system as a whole, learning implies the continual establishment of new interconnecting neurone pathways. Hence a continual increase in brain RNA with age is to be expected. The increase in any particular neurone will occur relatively rapidly, but the increase within any tissue mass will be slow.

It is clear that the administration of any compound to an animal that will interfere with the RNA-dependent biosyntheses in neurones will affect learning and memory. Azaguanine is such a compound. It is a structural analog of guanine (a purine base that is a major constituent of both DNA and RNA). It is to be predicted that the presence of this analog in brain cells will disrupt the inductive increase in enzymes, due to an inhibition of the DNA- and RNA-dependent processes necessary for enzyme biosynthesis.

The hypothesis predicts increases in those molecular species of RNA required for the biosynthesis of enzymes involved in the metabolism of nerve transmitter substances. Other types of RNA will not be altered. Now every molecular species of RNA probably has a different ratio of nitrogenous bases. Consequently, prior to induction the total RNA of a cell will have a particular base ratio that is the mean of all the individual RNA base ratios. After induction some types of RNA have increased. Hence the base ratio of total RNA will now be different.

It will be seen that the changes in RNA metabolism predicted by the enzyme induction hypothesis are just those that have been experimentally reported and claimed to support the nucleotide sequence hypothesis. It is clear that this evidence cannot be used to decide between these two alternate hypotheses.

One further piece of experimental evidence will be considered. It has been reported by Corning and John (1961) that planarians conditioned to avoid light can be cut in two and that after regeneration both halves retain the conditional response. However, if the planarians are allowed to regenerate in water containing small amounts of ribonuclease (an enzyme catalyzing the hydrolysis of RNA), only organisms regenerated from "heads" retain the response.

These experiments are difficult to interpret on several grounds. First, ribonuclease is a protein and may not diffuse into the planarians' cells, or may not diffuse equally into all. Second, there is no evidence as to the quantity and distribution of neurones involved in the response. Thus on the enzyme induction hypothesis, assuming that the response involves some interaction between the head ganglia and the somatic neurones, certain levels of RNA are necessary to maintain induced levels of transmitter substances in particular neurone pathways. On treating the "tail" with ribonuclease these essential RNA levels are reduced to the preinduction concentrations. Hence the organism regenerated from this portion lacks the conditioned response. On either hypothesis, it is necessary for the ribonuclease not to destroy the neurone RNA in the head section. Consequently, this experiment also fails to provide an experimental test to distinguish between the two hypotheses.

In summary it may be said that the nucleotide rearrangement hypothesis of memory postulates undemonstrated properties of nucleic acids that seem incompatible with current knowledge, while the hypothesis fails to give any satisfactory account of how RNA-stored information could influence the nervous mechanisms of memory. The experimental evidence produced in favor of this hypothesis is equally compatible with the hypothesis that memory is basically dependent upon neurone pathways maintained by high levels of transmitter substances due to induced biosynthetic enzymes. This enzyme induction hypothesis is also compatible with reported age and learning increases in brain cholinesterase levels: a phenomenon not predicted by the nucleotide rearrangement hypothesis.

REFERENCES

ABRAMS, R. Nucleic acid metabolism and biosynthesis. *Annu. Rev. Biochem.*, 1961, **30**, 165–188.

APRISON, M. H., & HIMWICH, H. E. Relation between age and cholinesterase activity in several rabbit brain areas. *Amer. J. Physiol.*, 1954, **179**, 502–506.

BENNETT, E. L., KRECH, D., ROSENZWEIG, M. R., KARLSSON, H., DYE, N., & OHLANDER, A. Cholinesterase and lactic dehydrogenase activity in the rat brain. *J. Neurochem.*, 1958, **3**, 153–160.

BENNETT, E. L., ROSENZWEIG, M. R., KRECH, D. R., KARLSSON, H., DYE, N., & OHLANDER, A. Individual, strain and age differences in cholinesterase activity of the rat brain. *J. Neurochem.*, 1958, **3**, 114–152.

CAMERON, D. E., & SOLYOM, L. Effects of ribonucleic acid on memory. *Geriatrics*, 1961, **16**, 74–81.

CORNING, W. C., & JOHN, E. R. Effect of ribonuclease on retention of conditioned response in regenerated planarians. *Science*, 1961, **134**, 1363–1365.

DINGMAN, W., & SPORN, M. B. The incorporation of B-azaguanine into rat brain RNA and its effect on maze learning by the rat. An inquiry into the biochemical basis of memory. *J. psychiat. Res*, 1961, **1**, 1–11.

DIXON, M., & WEBB, E. C. *Enzymes.* London: Longmans, 1958.

EGYHAZI, E., & HYDEN, H. Experimentally induced changes in the base composition of the ribonucleic acids of isolated nerve cells and their oligodendroglial cells. *J. biophys. biochem. Cytol.*, 1961, **10**, 403–410.

FINCHAM, J. R. S. Genetically controlled differences in enzyme activity. *Advanc. Enzymol.*, 1960, **22**, 1–43.

FREESE, A. B. Transitions and transversions induced by depurinating agents. *Proc. Nat. Acad. Sci., Wash.*, 1961, **47**, 540–545.

GAITO, J. A biochemical approach to learning and memory. *Psychol. Rev.*, 1961, **68**, 288.

HYDEN, H. Biochemical changes in glial cells and nerve cells at varying activity. In, *Proceedings of the international congress of biochemistry: Biochemistry of the central nervous system.* Vol. III. London: Pergamon, 1959, Pp. 64–89.

HYDEN, H. Satellite cells in the nervous system. *Scient. Amer.*, 1961, **205(6)**, 62–70.

KORNBERG, A. Biologic synthesis of deoxyribonucleic acids. *Science*, 1960, **131**, 1503–1508.

KRECH, D., ROSENZWEIG, M. R., & BENNETT, E. L. Dimensions of discrimination and the level of cholinesterase activity in the cerebral cortex of the rat. *J. comp. physiol. Psychol.*, 1956, **49**, 261–268.

KRECH, D., ROSENZWEIG, M. R., & BENNETT, E. L. Correlation between brain cholinesterase and brain weight within two strains of rats. *Amer. J. Physiol.*, 1959, **196**, 31–32.

KRECH, D., ROSENZWEIG, M. R., BENNETT, E. L., & LONGUEIL, C. L. Changes in brain chemistry of the rat following experience. Paper read at American Psychological Association, Cincinnati, September 1959.

LESLIE, I. Biochemistry of heredity: A general hypothesis. *Nature*, 1961, **189**, 260–268.

SPEYER, J. F., LENGYEL, P., BASILIO, C., & OCHOA, S. Synthetic polynucleotides and the amino acid code. *Proc. Nat. Acad. Sci., Washington*, 1962, **48**, 63–68.

SUEOKA, N. Correlation between base composition of DNA and amino acid composition of protein. *Proc. Nat. Acad. Sci., Wash.*, 1961, **47**, 1141–1149.

WEISS, S. A., & NAKAMOTO, T. On the participation of DNA in RNA biosynthesis. *Proc. Nat. Acad. Sci., Washington*, 1961, **47**, 694–697.

(Received December 11, 1961)

Psychological Review
1962, Vol. 69, No. 6, 542–547

THE UNTENABILITY OF LUCE'S PRINCIPLE

WILLIAM W. ROZEBOOM

Wesleyan University

In the course of his instructive researches in the mathematics of psychophysical scaling, R. Duncan Luce (1959a; 1959b, p. 29) has recently proposed a methodological principle which, if true, would have revolutionary import not merely for psychophysics, but for empirical and theoretical research in all branches of science. Moreover, while Luce's Principle has met with something less than enthusiastic acceptance in all quarters (cf. Luce, 1959b, p. 90; Mosteller, 1958, p. 287), others have begun to draw upon it as a basis for their own work (cf. Galanter & Messick, 1961). Such applications are decidedly premature, however, for the Principle's methodological status is dubious at best. As is inevitable for a provocatively novel hypothesis, certain ambiguities and unresolved difficulties reside in Luce's initial account, and when these are made explicit, it becomes highly problematic whether an interpretation can be found for the Principle which escapes being either vacuously true or empirically false.

Operating within the context of Stevens' theory of scale-types (e.g., Stevens, 1951),[1] Luce has proposed that the class of transformations defined by a scale-type should impose limitations on the mathematical form of the functional relations in which a variable of that type can participate. Specifically, it is postulated (Luce, 1959b) that:

> A substantive theory [i.e., an assumed law] relating two or more variables and the measurement theories [i.e., assumptions about the scale-types which are appropriate] for these variables should be such that:
>
> 1. (*Consistency of substantive and measurement theories*) Admissible transformations [i.e., transformations belonging to the class defined by the scale-type] of one or more of the independent variables shall lead, via the substantive theory, only to admissible transformations of the dependent variables.
>
> 2. (*Invariance of the substantive theory*) Except for the numerical values of parameters that reflect the effect on the dependent variables of admissible transformations of the independent variables, the mathematical structure of the substantive theory shall be independent of admissible transformations of the independent variables (p. 85).

Before putting his Principle to work, Luce comments that "one can hardly question the consistency part of the principle," but adds that "the invariance part is more subtle and controversial." I shall propose, to the contrary, that if this verbal statement of the Principle is taken at face value, it is the "consistency" part which is strong and correspondingly untenable, whereas the "invariance" part is a weakened version of the former which need be no more than a mathematical truism.

What can be meant when, in the consistency part of his Principle, Luce speaks of a transformation of an independent variable as *leading to*, via the substantive theory, a transformation of the dependent variable? The wording suggests that the transformation of the one somehow compels or induces a transformation of the other, an interpretation which is strengthened when, in subsequent illustration, Luce refers to a transformation being "effected on" the dependent variable. This cannot be a mathematical compulsion, however, for there is nothing about a lawful relationship which requires us to rescale one variable when transforming another. Thus given the law, "The area of a circle in square inches is π times its squared radius in inches," if I elect to scale length in feet rather than inches, my law simply becomes "The area of a circle in square inches is 144π times its squared radius in feet." (If in this example the rescaling of length is felt—as it should not be—to require a corresponding rescaling of area, we could instead take, e.g., the relation between a person's height and his age.) The apparent meaning of the "consistency" stricture comes clearer upon examination of Luce's proffered illustration:

---

[1] I should also like to voice my reservations about a number of the conclusions of this theory. However, the substance of these doubts must await another occasion.

> Suppose it is claimed that two ratio scales are related by a logarithmic law. An admissible transformation of the independent variable x is multiplication by a positive constant k, i.e., a change of unit. However, the fact that $\log kx = \log k + \log x$ means that an inadmissible transformation, namely, a change of zero, is effected on the dependent variable. *Hence the logarithm fails to meet the consistency requirement* [italics added] (1959b, p. 85).

So worded, the example does not make explicit what happens to the dependent variable; however, let y be a function of x by the relation $y = \log x$, and let x' ($=_{\text{def}} kx$) be the transformed scale of the independent variable. Then the cited mathematical fact about $\log kx$ implies that $\log x' = \log k + \log x$, or $y + \log k = \log x'$; so if we wish to find a rescaling, y', of the dependent variable which stands in the same relation to x' as that in which y stands to x, it has to be $y' = y + \log k$, which is inadmissible (says Stevens) for a ratio scale. As demonstrated by this example, therefore, what the consistency injunction entails is that if an admissible transformation is made of the independent variable, then there must be an admissible rescaling of the dependent variable which leaves the mathematical expression of the law invariant—i.e., if $y = \phi x$, and x' is an admissible rescaling of x, then there is an admissible rescaling, y', of y, such that $y' = \phi x'$. Other than this, it is difficult to see what sense could be made of the notion that transformation of the one variable "effects" a transformation of the other.

Before defending this interpretation against another which also needs to be considered, it is useful to state explicitly just what does happen to a lawful relationship, $y = \phi x$, under transformation of its variables.[2] Let T and U be identity preserving (i.e., one-one) transformations, admissible or otherwise, applicable to x and y, respectively. Then if x is transformed into x' by T and y is transformed into y' by U, we have $x' = Tx$ and $y' = Uy$, or $x = T^{-1}x'$ and $y = U^{-1}y'$. Hence the substantive theory

$$y = \phi x$$

rewritten in terms of x' and y', becomes

$$U^{-1}y' = T^{-1}x'$$

or

$$y' = U\phi T^{-1}x' \qquad [1]$$

Moreover, if it is *also* the case that $y' = \phi x'$, it follows from Equation 1 that $\phi x' = U\phi T^{-1}x'$, or $\phi Tx = U\phi x$. Let T_a be an *admissible* transformation of x, and U_a, similarly, an admissible transformation of y. Then the consistency claim of Luce's Principle apparently requires that if $y = \phi x$, the function ϕ must be such that for any admissible transformation T_a, there corresponds an admissible transformation U_a such that for all values of x,

$$\phi T_a x = U_a \phi x \qquad [2]$$

(It is easily seen that if the inverse of ϕ exists, this condition on ϕ is satisfied if and only if $\phi T_a \phi^{-1}$ is an admissible transformation of y.)

It has been proposed to me that what Luce intended the consistency part of his Principle to assert is only that if an admissible transformation, $x' = T_a x$, is made of the independent variable in $y = \phi x$, then there must exist some relationship ϕ' and an admissible transformation, $y' = U_a y$, of the dependent variable such that $y' = \phi' x'$. Then to derive Equation 2, the invariance part of the Principle must be construed to stipulate that U_a can always be chosen to set $\phi' = \phi$. But if this is all that the consistency claim were to imply, then it would have no restrictive force whatsoever—we can always take $\phi' = \phi T^{-1}$, in which case y requires no transformation at all—and hence a case could never arise, contrary to Luce's example cited earlier, in which the consistency requirement is violated. Moreover, as the invariance part of the Principle is worded, it is not strong enough to authorize Equation 2 because it insists not on *strict* invariance of ϕ under admissible transformations of the variables, but only on an invariance of some *structure* of ϕ. But the mathematical form (i.e., structure) of a function is hopelessly nonunique, and unless additional, as yet unspecified, conditions are imposed on this to-be-invariant form, we may take it to be a form common to any function of the class $\{U_a \phi T_a^{-1}\}$—which then leaves the invariance condition without any bite.

---

[2] Since x may be construed as a vector, the analysis which follows applies to multivariate relationships as well as to laws involving only one independent variable.

In any event, that Condition 2 is, in fact, the mathematical nugget which Luce (1959b) extracts from the ore of his Principle is demonstrated by his Table 1. Each entry in the "Functional Equation" column may be seen to be an instance of Equation 2. In view of the stiflingly strong consequences (e.g., Luce, 1959a, Table 2) of Luce's Principle, what can be said to justify it? Despite the undeniable mathematical charm of Equation 2, it seems to me that the answer must be an uncompromising "nothing whatsoever." It will be noticed that Luce himself does not actually present any argument to support his Principle, but in effect merely invites the assent of the reader's intuition. But intuition may well be gulled by the ambiguities in the verbal statement of the Principle into acceding to an innocuous reading of it which then imperceptibly slides into a not-so-innocuous version. This is especially likely for the consistency part of the Principle, which Luce insists can hardly be questioned even while at the same time holding it to imply at least *some* restriction on the form of a law. (But it has already been argued that the consistency claim is either vacuous or carries the full burden of Equation 2 by itself.) In any case, in my own methodological research I have seen too many initially self-evident theses wither away under careful probing to have much confidence in methodological intuition.

Pending cognitive support for Luce's Principle, then, how well does it stand the test of applicability to established laws? Well, as Luce himself points out, it is not difficult to find laws (see example below) which do *not* conform to it. Luce's comments on this faux pas carry the discussion into a rather different problem of vast proportions, but before pursuing his lead, we may pause to raise an eyebrow at the methodological status of a "principle" which is not binding. To be sure, the verbal statement of Luce's Principle does not say that all substantive laws *do* satisfy the Principle, but only that they *should* do so. It is most unlikely, however, that what was intended by this wording is merely that conformity to the Principle is *desirable*. Science is concerned with what is, not with what ought to be, and if the Principle is to have any scientific significance, it must be that Condition 2 says something about the forms which laws are *in fact* able to assume, and in light of which we may limit the forms to be considered when speculating about laws hitherto undiscovered. But if there are perfectly good laws which violate Equation 2, then it is difficult to see how the Principle could ever by itself impugn the legitimacy of any assumption about the form of a problematic law.

If Luce's Principle is to have any methodological force at all, therefore, there must exist additional criteria by which we can discriminate between laws which fall under the Principle's sovereignty and those which do not. Luce (1959b) has tentatively identified such a criterion: "All physical examples which have been suggested to me as counterexamples . . . have a common form: the independent variable is a ratio scale, but it enters the equation in a dimensionless fashion" (pp. 90 ff.). That is, the law can be put into a form in which the independent variable, x, is multiplied times a parameter, c, whose numerical value is determined by the choice of scale for x, so that any change in unit for x may be absorbed into c without disturbing the rest of the law—i.e., the law can be written in form $y = \phi_c(x) = \psi(cx)$, which under transformation $x' = kx$, goes over into $y = \phi_{c'}(x')$, where $c' = c/k$. Luce suggests that we can avoid violation of Condition 2 in such cases by taking the "dimensionless" variable v ($=_{\text{def}} cx$) as our independent variable; then the law becomes $y = \psi(v)$, in which v has no admissible transformations and there are hence no restrictions imposed on ψ by Condition 2. Thus a more subtle version of Luce's Principle emerges: What is now postulated is not that Condition 2 is inviolable, but that the laws of nature are always such as to be *capable* of an expression which satisfies Equation 2.

Clearly crucial to this revised formulation of Luce's Principle, however, is the concept of a dimensionless variable. For if the method for satisfying Equation 2 in prima facie embarrassing cases is to search out a dimensionless expression of the independent variables, the restrictive force (if any) of the Principle cannot be comprehended until we understand what is involved in such a construction. So pregnant a topic as this amply merits discussion and controversy, which I will do my bit to provoke by injecting some irritants which, hopefully, may stimulate Luce and perhaps others to further production of methodological pearls.

It is my considered opinion that to render a variable dimensionless in the fashion brought to our attention by Luce amounts to nothing but a more or less arbitrary selection of one of the admissible scalings of that variable, and then working up an "absolute" interpretation for that scale. I contend that such an interpretation can be found for at least one admissible transformation of any variable, no matter what its scale type (in fact, I would argue that *every* scale can have an absolute interpretation forced upon it), and hence that there is no law on whose mathematical form the revised version of Luce's Principle could impose restrictions. To defend this thesis in full generality is impractical here, so I shall confine my remarks to ratio scales.

An example cited by Luce of a law which flouts Condition 2, but which can be brought to heel by dedimensionalizing it, is the law of radioactive decay. If q is the quantity, say in grams, of a certain radioactive substance S at any time t, say the number of seconds after the beginning of the Twentieth century, then

$$q = ae^{-bt} \qquad [3]$$

where a and b are empirical constants. So formulated, the dependent variable (q) is a ratio scale while the independent variable (t) is an interval scale, which apparently contravenes not only Condition 2 (cf. Luce, 1959b, Theorem 4), but also Luce's suggestion that the only violations of Equation 2 are ones in which the independent variables are ratio scales. However, it is debatable whether generalizations such as Equation 3, in which a time coordinate assumes the status of an independent variable, should be regarded as genuine laws (cf. Rozeboom, 1961, pp. 356f.). If t_0 is an arbitrary point in time on the t scale, and q_0 is the quantity (on the q scale) of S at time t_0, then Equation 3 may be rewritten as

$$q = q_0e^{-b(t-t_0)}$$

or letting $d =_{\text{def}} t - t_0$,

$$q = q_0e^{-bd} \qquad [4]$$

Equation 4, in which all the variables are ratio scales, relates the quantity (in grams) of substance S at the end of a given time interval jointly to the interval's duration (in seconds) and S's quantity (in grams) at the interval's beginning. We can also express Equation 4 as a law with only one independent variable by dividing both sides by q_0 and setting $p =_{\text{def}} q/q_0$. We then have

$$p = e^{-bd}, \qquad [5]$$

which shows what proportion of S at a given moment still remains d seconds later, and in which the dependent variable, p, is an absolute, or dimensionless, scale.

Neither Equation 4 nor Equation 5

yet satisfy Condition 2,[3] but if we now ask how long a "half-life," L_h, substance S has—i.e., L_h is the value of d for which $p = .5$ in Equation 5—we find that $L_h = (\ln 2)/b$, and hence that

$$q = (\tfrac{1}{2})^{d/L_h} \qquad [6]$$

Then letting d_h be the ratio of d to L_h—that is, if the time duration variable is rescaled with the half-life of S as the unit of measurement—we finally obtain

$$q = (\tfrac{1}{2})^{d_h} \qquad [7]$$

in which the independent variable is an absolute scale with no admissible transformations, and which is hence immune to any threat from Condition 2. This, then, is the sort of procedure which, suggests Luce, may be used to placate Condition 2 when an empirical law involving ratio scales (e.g., Equation 4) appears to disregard it.

It should be appreciated, however, that to describe a variable such as d_h in Equation 7 as dimensionless is somewhat misleading. The scale is just as much dependent upon a unit of measurement as it ever was; only now one particular unit has been singled out as a *preferred* unit, which determines an absolute scale of the variable in the sense (and merely in this sense) that there is one and only one scale which expresses each value of the variable as a multiple of this preferred unit. This can obviously be done whenever a preferred unit of measurement has been selected, and hence *any* ratio scale law, no matter what its mathematical structure, can in this fashion be put into a form which is compatible with Equation 2.

There is still one more gambit by which some restrictive force might conceivably be salvaged for Luce's Principle: If a distinction could be drawn between *natural* units of measurement on one hand, and merely conventional ones on the other, the definition of dimensionless scales could be limited to variables for which a natural unit is available. While it is my belief that "naturalness" in this sense is purely a psychological effect with little or no methodological significance, I will not argue the point here. Instead, I merely invite anyone who feels otherwise to state what unit is the "natural" one for dedimensionalizing the time-duration variable in the law of radioactive decay. There is nothing special about the half-life interval. For any proportion r, we could just as well take substance S's r-life, L_r, for our standard (i.e., L_r is the time required for S to decay to the proportion r of its initial amount), and write our law as

$$p = r^{d/L_r} \qquad [6']$$

or

$$p = r^{d_r}. \qquad [7']$$

where d_r is the L_r-scale of duration. Preference for $r = .5$ is wholly a matter of psychological convenience.

SUMMARY

The *strong* version of Luce's Principle, namely, that all laws must conform to Condition 2, is not only unsupported by reason, but is refuted by counterexamples. On the other hand, a modified version of the Principle which demands of a law only that it be equivalent to one which satisfies Equation 2 (or, somewhat less generally, that a law is exempt from Equation 2 only if its independent variables can be dedimensionalized) is vacuous insomuch as any law, no matter how seriously at odds with Equation 2 in its original form, can be coaxed into compliance with the weakened requirement. Until further clarifications and supports for Luce's Principle are forthcoming, then, it would be questionable strategy to limit speculations about possible laws to mathematical forms approved by Equation 2.

---

[3] Equation 4 violates Luce's Theorem 1 (1959b), while Equation 5 is especially interesting in that since its inverse exists and its dependent variable has no admissible transformations (other than the Identity transformation), Condition 2 requires that the same be true of its independent variable.

REFERENCES

GALANTER, E., & MESSICK, S. The relation between category and magnitude scales of loudness. *Psychol. Rev.*, 1961, **68**, 363–372.

LUCE, R. D. *Individual choice behavior: A theoretical analysis.* New York: Wiley, 1959. (a)

LUCE, R. D. On the possible psychophysical laws. *Psychol. Rev.*, 1959, **66**, 81–95. (b)

MOSTELLER, F. The mystery of the missing corpus. *Psychometrika*, 1958, **23**, 279–289.

ROZEBOOM, W. W. Ontological induction and the logical typology of scientific variables. *Phil. Sci.*, 1961, **28**, 337–377.

STEVENS, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1–49.

(Received February 14, 1962)

Psychological Review
1962, Vol. 69, No. 6, 548–551

COMMENTS ON ROZEBOOM'S CRITICISMS OF "ON THE POSSIBLE PSYCHOPHYSICAL LAWS"[1]

R. DUNCAN LUCE

University of Pennsylvania

The publication of Rozeboom's (1962) criticisms of my paper concerning limitations on scientific laws of certain types (Luce, 1959), together with what we both feel are misinterpretations and misleading uses of those results, prompt me to attempt to clarify some of the issues. At the outset, let me admit that much of the confusion is my fault, due in part to an inadequate understanding of the problem and in part to a not wholly satisfactory presentation of the ideas and results. These failures have been aggravated by some readers who seem to have ignored the second half of that paper and, thereby, have oversimplified its content.

Rozeboom takes me to task for an ambiguous statement of the two criteria that were suggested as a possible "principle of theory construction"—despite these ambiguities, which he has discussed in detail, we both agree about the mathematical "nugget" to be extracted from them, namely his Equation 2. We both emphasize that there are physical laws that either fail to satisfy this equation or satisfy it vacuously, and he thus concludes, correctly I now believe, that it should never have been termed a "principle." We also both point out that whether Equation 2 has any bite in limiting the form of a law has something to do with the existence of dimensional parameters in the law. We seem to differ mainly on the emphasis to place upon this distinction. Because I believe that it is extremely important, I should like to make a few remarks about it.

Among the various lawful statements that can be found in any science, Rozeboom and I are restricting our attention to what may be called function laws—the assertion of a functional relation among several variables, each of which can be represented as a simple numerical scale with certain invariance properties dictated by its measurement theory. As a specific example, from which we will generalize, consider the familiar law $E = mc^2$, where E denotes energy, m denotes mass, and c the velocity of light in a vacuum, all measured in some consistent set of units. It will be convenient to think of this law in the form $E - mc^2 = 0$. The first thing involved in this physical law is the assertion that there exists a particular function of three variables, call it ψ, such that all physically realizable values of E, m, and c satisfy

$$\psi(E, m, c) = 0$$

But there is a good deal more intended than just this simple mathematical equation. For example, not all of the arguments in the equation are of the same type: they are constrained in different ways. We usually speak of E as the dependent variable, m as the independent variable, and c as a dimensional parameter. By these words we mean such things as this: that certain of the symbols can assume arbitrary (positive) values and that the law then constrains others to have particular values, that other symbols represent constants that cannot be experimentally manipulated, and that certain measurement transformations of some of the symbols can be carried out provided that the transformations are admissible according to the measurement theories for these variables and that certain variables and parameters are transformed in prescribed ways. In other words, the statement of a law involves not only the usually written equation, but also certain specifications about which values can be arbitrarily selected, about which variables can be freely transformed, and about which values and transformations are prescribed when these choices are made.

---

[1] This work was supported in part by National Science Foundation Grant G-17637 to the University of Pennsylvania.

To generalize from the example, we distinguish three kinds of symbols that enter into our functional relation (the distinctions among them are rendered more precise later): There are $m + 1$ variables, one of which is designated as

1. the dependent variable, y, and it assumes values in a set Y of real numbers, and the other $m \geqslant 1$ are designated as
2. independent variables, x_i, and they assume values in, respectively, sets X_i of real numbers; in addition, there are
3. $n \geqslant 0$ dimensional parameters, a_j.

Associated with each of these quantities is a group of admissible transformations. In the case of a variable, the group is determined by its measurement theory whereas, in the case of a parameter, it is simply included as part of the statement of the law. We denote these groups by $\mathbf{Y}$, $\mathbf{X}_i$, and $\mathbf{A}_i$. The more common transformation groups found in science include the affine group (multiplication by a positive constant), the full linear group, and the group of strictly monotonic increasing transformations—corresponding, respectively, to what are called ratio, interval, and ordinal scales.

Once the variables, parameters, and transformation groups are listed, the statement of a law then involves two distinct assertions. First, a particular function ψ of $1 + m + n$ variables is given which has the property that for any physically realizable choice of $x_i \epsilon X_i$, $i = 1, 2, \cdots, m$, any physically realizable value $y(x_1, \cdots, x_m) \epsilon Y$ of the dependent variable satisfies

$$\psi[y(x_1, \cdots, x_m), x_1, \cdots, x_m, a_1 \cdots, a_n] = 0$$

Second, for any set of (admissible) transformations $T_i \epsilon \mathbf{X}_i$, $i = 1, 2, \cdots, m$, of the independent variables, there exists an (admissible) transformation $U(T_1, \cdots, T_m) \epsilon \mathbf{Y}$ of the dependent variable and a set of (admissible) transformations $S_j(T_1, \cdots, T_m) \epsilon \mathbf{A}_j$, $j = 1, 2, \cdots, n$, of the dimensional parameters such that

$$\psi(y, x_1, \cdots, x_m, a_1, \cdots, a_n) = 0$$

implies

$$\psi[U(T_1, \cdots, T_m)y, T_1x_1, \cdots, T_mx_m, S_1(T_1, \cdots, T_m)a_1, \cdots, S_n(T_1, \cdots, T_m)a_n] = 0.$$

The first part of such a law, the function relating the variables, is always stated explicitly, but the second part is often implicit. It is simply a convention that scientists know and take for granted. For example, if a decay law for a particular radioactive material is

$$q = q_0e^{-0.14t}$$

when the time t since a certain event is measured in seconds, then everyone knows that if we measure time in hours we must change the time constant from 0.14 to (0.14)(360) = 50.4.

Despite the fact that they are often left implicit, these transformation properties are important and must be made explicit in a full statement of a law. Indeed, one of the things that distinguishes theoretical science from a collection of simple numerical assertions are the relations among admissible transformations. In addition, the relations among transformations make the distinctions among the arguments of a law more than just linguistic conventions. Note that the independent variables are the arguments of ψ whose values in the first part and whose transformations in the second are selected arbitrarily; the value and transformation of the dependent variable are both determined via the law by those of the independent variable; and the value of a dimensional parameter is independent of the values of the variables and its transformation is determined by those applied to the independent variables. It is often possible to restate the law so that a different variable is treated as dependent; but this is not always possible, as for example, when the dependent variable is a periodic function of the independent ones.

Now consider a law (such as $F = ma$) that includes no dimensional parameters, and suppose that the solution y of

$$\psi(y, x_1, \cdots, x_m) = 0$$

is unique, i.e., there is a function ϕ of m variables such that

$$y = \phi(x_1, \cdots, x_m)$$

solves the above equation. Then the second part of the law can be rewritten as: for any $T_i \epsilon \mathbf{X}_i$, $i = 1, 2, \cdots, m$, then there exists $U(T_1, \cdots, T_m) \epsilon \mathbf{Y}$ such that if

$$y = \phi(x_1, \cdots, x_m)$$

then

$$U(T_1, \cdots, T_m)y = \phi(T_1x_1, \cdots, T_mx_m)$$

Substituting the first of these equations into the second, we obtain the following restriction on the function

$$U(T_1, \cdots, T_m)\phi(x_1, \cdots, x_m) = \phi(T_1x_1, \cdots, T_mx_m)$$

which is Rozeboom's Equation 2 for m independent variables; it includes the several special cases investigated in Luce (1959). For ratio and interval scales and $m = 1$, it was shown that, aside from numerical constants, there are very few possible solutions ϕ to this functional equation.

When dimensional parameters are admitted, the resulting equation is very much less restrictive. Specifically, with one independent variable and one dimensional parameter, the same argument leads to

$$U(T)\phi(x, a) = \phi[Tx, S(T)a]$$

which has many possible solutions. For example, suppose that both $y = \phi(x, a)$ and x are variables measured on ratio scales, and suppose that f is an arbitrary function of one variable. A solution to this equation is $\phi(x, a) = f(xa)$ provided that when x is transformed by $Tx = kx$, then U is the identity transformation and $S = T^{-1}$, i.e., $Uy = y$ and $Sa = a/k$.

Two conclusions seem justified. First, Rozeboom is correct in saying that the equation that results when there are no dimensional parameters is much too restrictive to be called a principle of theory construction. Second, and this I feel to be the main thrust of both my 1959 paper and this note, psychologists (as well as other scientists) either are restricted to a very few possible types of laws (in the case of one independent variable, to those given in my 1959 paper when the variables are measured on either ratio or interval scales) or they cannot avoid including dimensional parameters in the statement of their laws. Not only must these parameters be written explicitly, but a careful prescription should be included about how they transform as a function of admissible transformations of the independent variables, for otherwise the statement of the law is incomplete. Following the practice of physicists, the transformational aspects of laws are often not made explicit. This does not seem to create serious problems when the transformations are limited to the affine group (ratio scales), but matters are much less obvious when interval and weaker scales are involved. In these cases, which often seem to arise in psychology, it is essential that careful attention be given to the transformation properties of the variables and parameters.

Finally, the problem remains whether or not the number of dimensional parameters involved in the statement of a law is of any inherent importance. Some physicists seem to feel that such parameters are undesirable and that their total number in a science should be held to a minimum. My earlier results strongly suggest that the minimum cannot possibly be zero. Although psychologists have hardly begun to face these problems, inevitably they will. At that time, it would be useful to have a better understanding of the role of dimensional parameters than we have now. In physics they have meaning—the velocity of light, the density of a substance, the viscosity of a fluid, the gravitational constant, etc.—and they recur in various combinations in different laws. As yet, we have little of this in psychology.

SUMMARY

It is pointed out that certain scientific laws assert not only a functional relation among variables and dimensional parameters, but that they also state what transformations on the dependent variable and the parameters are necessary for the same function to be satisfied when admissible transformations are applied to the independent variables. It is shown when there are no dimensional parameters, these transformation conditions are very restrictive but that they are not when one or more dimensional parameters are present. The conclusion is that psychological laws will in general involve dimensional parameters and that it is necessary to state carefully the transformational aspects of the law.

REFERENCES

LUCE, R. D. On the possible psychophysical laws. *Psychol. Rev.*, 1959, **66**, 81–95.

ROZEBOOM, W. W. The untenability of Luce's principal. *Psychol. Rev.*, 1962, **69**, 542–547.

(Received May 9, 1962)

Psychological Review
1962, Vol. 69, No. 6, 552

COMMENT

WILLIAM W. ROZEBOOM

Wesleyan University

I am grateful to the Editor for allowing me this opportunity to express my pleasure at Luce's gracious acceptance of my criticisms of his earlier paper, and to record my full endorsement of his present conclusions. Briefly, the point is simply this: If a law is stated without stipulation of one particular scale for each of the variables involved, and if, moreover, it is intended that alternative scaling of these variables be countenanced, then it is also necessary to make explicit provision in the law's statement for the manner in which the contemplated transformations will affect it. In particular, if it is assumed that the function relating the variables is invariant under these transformations—and I think that Luce is quite right in his judgment that this assumption is normally made, though I am much less convinced of its methodological necessity—then it will usually be necessary to include scale parameters in the law whose meanings and transformational behavior also require study and clarification. Thanks to Luce, explicit awareness of this need has now been entered in the mainstream of psychological literature, and a few more bolts on psychology's shaky methodological scaffolding have begun to tighten.

(Received June 8, 1962)

ERRATA

In the article by Frank Restle entitled "The Selection of Strategies in Cue Learning" (*Psychol. Rev.*, 1962, **69**, 329–343), the first equation at the top of page 333 should read:

$$\cdot \frac{N_c + N_i - N_{n,1} - \cdots - N_{n,j-2} - N_{n,j-1}}{N_c + N_i - N_{n,1} - \cdots - N_{n,j-3} - N_{n,j-2}}$$

$$\cdot \frac{N_{n,j}}{N_c + N_i - N_{n,2} - \cdots - N_{n,j-1} - N_{j-1}}$$

In the Errata on page 378 of the same issue, a line was omitted. It should read as follows (the omitted line appearing in italics):

"In the table of contents on the front cover of the May 1962 issue, *the title of the article by Joan E. Jones should read as follows:* '*Con*tiguity and Reinforcement in Relation to CS-UCS Intervals in Classical Aversive Conditioning'."